中国多世代人口数据库——辽宁部分（CMGPD-LN）

使用指南
中文翻本版 1.0（英文版本3.0为基础）

若引用本指南的内容，推荐的引用地址为：

Lee, James Z, Cameron Campbell, and Shuang Chen. 2010. *China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN) 1749-1909. User Guide.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

若使用中国多世代人口数据库——辽宁部分（CMGPD-LN），推荐的引用地址为：

Lee, James Z., and Cameron D. Campbell. China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909 [Computer file]. ICPSR27063-v6. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-09-02. doi:10.3886/ICPSR27063.v6.

鸣谢支持

CMGPD-LN的准备工作和通过ICPSR DSDR的公共文件发行接受American Recovery and Reinvestment Act的资金支持，立项编号为NICHD R01 HD057175-01A1 "Multi-Generation Family and Life History Panel Dataset"。

如果您想在著作中引用本指南或者数据库，请将完整的文献发送至cmgpd@icpsr.umich.edu ，以便我们可以将它补充进数据库的应用成果报告中。

若有问题、评论或建议，欢迎发送至 cmgpd@icpsr.umich.edu, 或者登录 **http://www.icpsr.umich.edu/icpsrweb/CMGPD/contact.jsp** 填写在线反馈表格。

目录

目录 2
表格目录 8
地图目录 10
鸣谢 12
引言 15
1 辽东八旗户口册 19
1.A 盛京内务府 20
1.B 盛京内务府与八旗人口登记 22
1.C 从户口册到中国多世代人口数据库——辽宁部分（CMGPD-LN） 24
2 使用 CMGPD-LN 28
2.A 概述 28
2.B 特点 28
2.B.i 前瞻性 28
2.B.ii 历时性 29
2.B.iii 封闭性 29
2.B.iv 详尽性 29
2.B.v 多层次性 30
2.B.vi 多世代 30
2.C 局限性 32
2.C.i 儿童遗漏 32
2.C.ii 遗失的户口册 32
2.C.iii 事件的发生时间 33
2.C.iv 覆盖面与代表性 33
2.D 应用实例 34
2.D.i 死亡率 34
2.D.ii 健康状况 35
2.D.iii 生育 35
2.D.iv 迁移 36
2.D.v 社会经济地位 36
2.D.vi 家庭，亲属关系与社区 37
2.D.vii 空间分析 37
2.D.viii 制度背景 37
3 CMGPD-LN 中基本变量说明 39
3.A 组织 39
3.B 处理缺失值 39
3.C 基本版数据（DS1） 39
3.C.i 原始变量 39
3.C.i.a 基本变量 42
DATASET 人群数据 42

YEAR 年份.....43
RELATIONSHIP 关系.....44
DISABILITY_CODE 残疾代码.....45
POSITION_CODE 地位代码.....45
POSITION_2_CODE 地位代码 2.....46
SEX 性别.....46
GENERATION 世代.....47
MARITAL_STATUS 婚姻状况.....47
AGE_IN_SUI 年龄（岁）.....48
BIRTHYEAR 出生年月.....49
BAZI_YEAR 生辰八字年份.....49
BAZI_MONTH 生辰八字月份.....50
BAZI_DAY 生辰八字日期.....50
BAZI_HOUR 生辰八字时辰.....50
3.C.i.b 居住地.....50
REGION（区域）.....50
DISTRICT 行政区划.....51
UNIQUE_VILLAGE_ID（唯一）村庄名称.....51
3.C.i.c 人口学事件.....52
DIED 死亡.....52
MARRIED OUT 外嫁.....53
REMARRIED_OUT 再次外嫁.....53
ABSCONDED 逃.....54
PRESENT 存活.....55
3.C.i.d 行政地位.....55
RETIRED 退.....55
OLD 老.....56
HAS_POSITION 在职.....56
ARTISAN 工匠.....57
SOLDIER 士兵.....58
NO _STATUS 无职位.....58
ZU ZHANG 族长.....59
3.C.ii 身份识别变量.....60
RECORD NUMBER （记录编号）.....60
PERSON ID 个人身份编号.....61
MOTHER_ID 母亲身份编号.....61
FATHER_ID 父亲身份编号.....61
FATHER_ID_IMPUTED 父亲身份假定编号.....62
GRANDFATHER ID 祖父身份编号.....62
GRANDFATHER ID IMPUTED（祖父身份假定编号）.....62
WIFE_1_ID 第一个妻子的身份信息.....62
WIFE_2_ID 第二个妻子的身份信息.....63
HUSBAND_ID 丈夫身份信息.....63
REGISTER_SEQ 登记序列号.....63

ZU_SEQ 族序列号 63
HOUSEHOLD_SEQ 户序列号 64
HOUSEHOLD_ID 户编号 64
3.C.iii 分析变量 64
3.C.iii.a 计数变量 64
BROTHER_COUNT 兄弟个数 64
SISTER_COUNT 姐妹个数 64
MALE_COUSIN_COUNT 堂兄弟个数 64
FEMALE_COUSIN_COUNT 堂姐妹个数 65
UNCLE_COUNT 叔伯个数 65
AUNT_COUNT 姑姑个数 65
FATHER_ALIVE 父亲在世 65
MOTHER_ALIVE 母亲在世 65
SON_COUNT 儿子数目 65
DAUGHTER_COUNT 女儿数目 65
3.C.iii.b 事件发生的标识变量 66
NEXT DIE 即将去世 66
NEXT_MARRY 即将婚嫁 67
NEXT_REMARRY 即将再婚 67
NEXT_ABSCONDED 即将逃 68
NEXT_BOYS 即将有男孩出生 68
NEXT_GIRLS 即将出生女孩 68
NEXT_3 下个三年 68
NEXT_6 下个六年 69
3.C.iii.c 确定相关事件的潜在人群的标识变量 69
AT_RISK_DIE 死亡风险 69
AT_RISK_MARRY 结婚风险 69
AT_RISK_REMARRY 再婚风险 69
EVER_ABSCONDED 曾经出逃 69
3.D 分析版数据（DS3） 70
3.D.i Positions and Statuses 职位和社会地位 70
BAI_ZONG 百总 70
BANNER 旗 70
BAOYANG 抱养 71
BI_TIE_SHI 笔帖式 71
DING_DAI 顶带 71
DISABLED 残疾的 71
ESTIMATED_INCOME 估计收入 73
EXAMINATION 考试 73
EXPELLED 开除 73
GAO_LI 高丽 74
DEAD 死亡 74
GUAN_XUE_SHENG 官学生 74
HONORIFIC 荣誉头衔 74

JUAN 捐.....75
MAN_ZHOU 满洲.....75
LING_CUI 领催.....75
NEW_DING 新丁.....75
RANK 等级.....76
PURCHASED_TITLE 购买的头衔.....76
QIAN_ZONG 千总.....76
SERVICE_DING 服丁.....76
TOU_CONG 投从.....77
XIAO_QI_XIAO 骁骑校.....78
ZHI_SHI_REN 执事人.....78
3.D.ii Family and Household 家庭与户.....78
BIRTH_ORDER 出生顺序.....78
BIRTH_ORDER_SEX 出生顺序_性别.....78
HH_SIZE 户的规模.....78
HH_DIVIDE_NEXT 下次户口册中家庭分割情况.....78
SENIORITY 年长.....78
SENIORITY_SEX 年长_性别.....79
3.D.iii Kin Groups 血亲组.....79
FOUNDER_ID 祖先身份编号.....79
FOUNDER_INFERRED_ID 祖先身份假定编号.....79
UNIQUE_GROUP 特殊一组.....80
UNIQUE_HH_ID 特殊户编号.....80
UNIQUE_YI_HU 特殊一户.....80
3.D.iv Events 事件.....81
ADOPTED_IN 领养.....81
ADOPTED_OUT 给别人收养.....81
EVENT_1 事件一.....81
EVENT_2 事件二.....81
3.D.v Naming 命名.....82
DIMINUTIVE_NAME 昵称.....83
HAS_SURNAME 有姓氏.....83
NON_HAN_NAME 非汉族名字.....84
NUMBER_NAME 数字名字.....85
RUSTIC_NAME 土名.....85
3.E Kinship File 血亲版数据(DS4).....85
3.E.i Ancestry Identifiers 祖先标识.....85
F_ID_1, F_ID_2, F_ID_3, F_ID_4 父亲、祖父、曾祖父、高祖父.....86
M_ID_1, M_ID_2, M_ID_3, M_ID_4 母亲，祖母，曾祖母，高祖母.....86
3.E.ii Kin Count Variables (Prefixed by KIN_) 亲属数量变量（前缀是 KIN_）.....87
3.F Restricted File 限制版数据(DS2).....88
RECORD_NUMBER 记录编号.....88
DATASET 数据人群.....88
RAW_VILLAGE_ID 原始村庄号码.....88

NAME 姓名 88
SURNAME_YIHU 一户姓氏 89
ASSIGNED_NAME 赋予的名字 89
LATITUDE 纬度 89
LONGITUDE 经度 89
SUOSHU 所属 90
4 CMGPD-LN Supplemental Files CMGPD-LN 补充文件 90
4.A Disability Supplement 残疾变量补充说明(DS5) 90
4.B Position Supplement 职位变量补充说明(DS6) 90
4.C Position 2 Supplement 职位变量补充说明 2 90
4.D Monthly Grain Prices 每月粮食价格 91
YEAR 年份 91
GRAIN_TYPE 粮食类型 92
HIGH_LOW 高低 92
MONTH01-MONTH12 月份 01-月份 12 92
4.E Annual Grain Prices 每年粮食价格 92
YEAR 年份 92
LOW_RICE, LOW_HUSKED_MILLET, LOW_UNHUSKED_MILLET, LOW_SORGHUM, LOW_WHEAT, LOW_SOY 低价大米、低价黍米、低价小米、低价高粱、低价小麦、低价大豆 93
HIGH_RICE, HIGH_HUSKED_MILLET, HIGH_UNHUSKED_MILLET, HIGH_SORGHUM, HIGH_WHEAT, HIGH_SOY 高价大米、高价黍米、高价小米、高价高粱、高价小麦、高价大豆 93
4.F Yingkou Customs Records 营口海关记录 93
YEAR 年份 93
EXPORT_BEAN 出口豆类 93
EXPORT_BEAN_CAKE 出口豆粕 93
EXPORT_BEAN_OIL 出口豆油 93
EXPORT_BEAN_ALL 出口豆类产品 93
IMPORT_VALUE 进口价值 93
EXPORT_VALUE 出口价值 94
TOTAL_VALUE 总价值 94
4.G Natural Disasters 自然灾害 94
YEAR 年份 94
DAY, HOUR and MINUTE 日期、时刻和分钟 95
EVENT_TYPE 事件类型 95
DISTRICTS_PINYIN 地区拼音 95
WATERSHEDS_PINYIN 分水岭拼音 95
EPICENTER_PINYIN 震中拼音 95
EPICENTER_LONGITUDE, EPICENTER_LATITUDE 震中经度、震中纬度 95
EARTHQUAKE_MAGNITUDE 地震强度 95
DISTRICT_01-DISTRICT_13 地区 01-地区 13 95
MONTH_01-MONTH_12 月份 01-月份 12 95

SEASON_1-SEASON_4 季节 01-季节 04.....95
5 中国多世代人口数据库-辽宁部分（CMGPD-LN）人口分类.....96
5.A 正身旗人户口册.....96
5.B 特殊义务旗人户口册.....98
5.C 庄园旗人户口册.....98
5.D 其他奴役人口.....100
5.E 其他特殊人口户口册.....100
6 中国多世代人口数据库——辽宁部分（CMGPD-LN）人口概况.....100
6.A 道义屯汉军人丁(1).....106
6.B 盖州汉军人丁(2).....108
6.C 打蜜汉军人丁(3).....109
6.D 城内汉军人丁(4).....111
6.E 盖州汉军棉丁(5).....112
6.F 牛庄刘二堡汉军人丁(6).....114
6.G 肥地一面城汉军人丁(7).....116
6.H 盖州满汉人丁(8).....118
6.I 大甸子汉军人丁(9).....119
6.J 郭三屯汉军人丁(10).....121
6.K 八棵树老古洞满汉人丁(11).....122
6.L 大辛屯汉军人丁(12).....124
6.M 年马岱海寨汉军人丁(13).....125
6.N 长寨子汉军人丁(14).....127
6.O 照化屯汉军人丁(15).....128
6.P 貂皮屯汉军人丁(16).....130
6.Q 郎家堡魏家屯汤岗子海州岳州汉军人丁(17).....132
6.R 王指挥屯汉军人丁(18).....133
6.S 阿尔吉善(19).....135
6.T 海城汉军人丁(20).....137
6.U 王多罗树满洲牲丁(21).....138
6.V 王多罗树满洲人丁(22).....140
6.W 瓦子峪汉军人丁(23).....142
6.X 五户汉军人丁(24).....143
6.Y 棉花盐靛庄头庄丁(25).....145
6.Z 苏拜公人丁(26).....147
6.AA 开档(27).....150
6.AB 开档投从抱养人丁(28).....151
6.AC 棉花盐靛小门人丁(29).....153

表格目录

表 1 CMGPD-LN 包含的 29 个数据库 15
表 2 RELATIONSHIP 编码解析 44
表 3 个人被列为死亡（故）的观察值数量 52
表 4 不同性别的外嫁记录 53
表 5 不同外嫁登记次数的人口数 53
表 6 男女再次外嫁的观测值数量 54
表 7 不同再次外嫁次数的丧偶人数 54
表 8 男女逃亡的观测值数量 55
表 9 退休人员年龄分布 55
表 10 个人第一次被标注为“老”时的年龄分布 56
表 11 拥有带俸禄职位的 18-60 岁男性分布 57
表 12 曾为工匠的 18-60 岁男性分布 57
表 13 曾为士兵的 18-60 岁男性分布 58
表 14 18-60 岁男性无职位分布 58
表 15 不同人口分类 20-50 岁成年男性残疾数量 72
表 16 作为官方职位修饰的荣誉头衔 73
表 17 不同数据人群中男性服丁分布 76
表 18 EVENT_1 和 EVENT_2 编码解释 81
表 19 前缀为 KIN_变量的名字组成 86
表 20 盛京内务府户口册中正身旗人、特殊义务旗人和庄园旗人比较 96
表 21 29 个行政单位的人口规模和地理分布 99
表 22 29 个行政区域人口的地区流动，按照每个个人居住的地区的数量 101
表 23 不同人群数据 18-26 岁男性中有官职、工匠、士兵的数量 103
表 24 道义屯汉军旗特征 106
表 25 盖州汉军旗特征 107
表 26 打蜜汉军旗特征 109
表 27 城内汉军人口特征 110
表 28 盖州汉军棉丁人口特征 112
表 29 牛庄刘二堡汉军人口特征 114
表 30 肥地一面城汉军人口特征 116
表 31 盖州满汉人丁特征 117
表 32 大甸子汉军村庄人口特征 119
表 33 郭三屯汉军人丁特征 120
表 34 八棵树老古洞满汉人丁特点 122
表 35 大辛屯汉军人丁特征 123
表 36 马岱海寨汉军人丁特征 125
表 37 长寨子汉军人丁特征 126
表 38 照化屯汉军人丁特征 128
表 39 貂皮屯汉军人丁特征 130
表 40 郎家堡魏家屯汤岗子海州岳州汉军人丁特征 132
表 41 王指挥屯汉军人丁特征 133

表 42 阿尔吉善人口特征.....135
表 43 海城汉军人丁特征.....136
表 44 王多罗树满洲牲丁特征.....138
表 45 王多罗树满洲人丁特征.....140
表 46 瓦子峪汉军人丁特征.....141
表 47 五户汉军人丁特征.....143
表 48 棉花盐靛庄头庄丁特征.....145
表 49 苏拜公人丁特征.....146
表 50 开档人口特征.....149
表 51 开档投从抱养人丁特征.....151
表 52 棉花盐靛小门人丁特征.....153

地图目录

地图 1 CMGPD-LN 中已经被定位的村落 18
地图 2 辽东地区示意图，对应 REGION 变量中的四个区划 51
地图 3 道义地区人口 106
地图 4 盖州汉军旗人口 107
地图 5 打蜜汉军旗村庄分布 109
地图 6 盖州汉军棉丁村庄分布 112
地图 7 牛庄刘二堡汉军村庄分布 114
地图 8 肥地一面城汉军人丁村庄分布 116
地图 9 盖州满汉人丁村庄分布 117
地图 10 大甸子汉军村庄分布 119
地图 11 郭三屯汉军人丁村庄分布 120
地图 12 八棵树老古洞满汉人丁村庄分布 122
地图 13 大辛屯汉军人丁村庄分布 123
地图 14 马岱海寨汉军人丁村庄分布 125
地图 15 长寨子汉军人丁村庄分布 126
地图 16 照化屯汉军人丁村庄分布 128
地图 17 貂皮屯汉军人丁村庄 130
地图 18 郎家堡魏家屯汤岗子海州岳州汉军人丁村庄分布 132
地图 19 王指挥屯汉军人丁村庄分布 133
地图 20 阿尔吉善村庄分布 135
地图 21 海城汉军人丁村庄分布 136
地图 22 王多罗树满洲牲丁村庄分布 138
地图 23 王多罗树满洲人丁村庄分布 140
地图 24 瓦子峪汉军人丁村庄分布 141
地图 25 五户汉军人丁村庄分布 143
地图 26 棉花盐靛庄头庄丁村庄分布 145
地图 27 苏拜公人丁村庄分布 146
地图 28 开档村庄分布 149
地图 29 开档投从抱养人丁村庄分布 151
地图 30 棉花盐靛小门人丁村庄特征 153

图片目录

图 1　肥地一面城 1783 年户口册式样.....24
图 2 肥地一面城 1792 年户口册式样.....24
图 3 图 2 中肥地一面城 1792 年户口册式样的数据录入.....25
图 4 肥地一面城 1795 年户口册前 2 页的数据录入.....26
图 5 不同时期儿童可以追溯到特定祖先的比例.....31
图 6 编纂于 1822 年的郎家堡魏家屯汤岗子海州岳州人丁户口册的封面.....41
图 7 辽东八旗户口册登记页式样.....42
图 8 不同 SEX 和 AGE_IN_SUI 的样本分布.....49
图 9 18-60 岁男性无行政地位的比例，1749-1909.....59
图 10 男女不同年龄未来三年的死亡概率.....67
图 11 21 岁以上编码为残疾的男性比例，以年为单位.....72
图 12 CMGPD_LN 中新丁的年龄分布.....75
图 13 不同年龄男性有昵称的比例.....82
图 14 不同年龄男性记录有姓氏的比例.....83
图 15 不同年份有非汉族名字的男性的比例.....84
图 16　1864 年苏拜公人丁户口册封面.....148

鸣谢

中国多世代人口数据库——辽宁部分（CMGPD-LN）是一部毕生之作，或者说是两个人：李中清先生（James Lee）和康文林先生（Cameron Campbell）的毕生之作。在这里，我们想要感谢许多个人和机构，在过去30年中，他们为数据库的创建和其内容的丰富做出了极大的贡献。

我们尤其受惠于鞠德源和第一历史档案馆，是他们最先提醒了李中清先生这些数据的存在。还有Melvin Thatcher和 犹他家谱学会（the Genealogical Society of Utah），是他们提供了户口册的影印版本供我们查阅，促成了中国多世代人口数据库 ——辽宁部分（CMGPD-LN）[1]的创建。我们还要感谢高静和辽宁省地方志办公室的工作人员，是他们陪伴我们进行实地考察，使得搜集CMGPD-LN地方版数据成为可能，基本版数据、分析版数据、血亲版数据和限制版数据发布后，我们将会发布这些地方版数据。辽宁省档案馆也高度配合和支持我们的工作，在李中清先生于1982和1985年，李中清和康文林先生于1987年的三次拜访期间，CMGPD-LN数据库1的大部分和CMGPD-LN数据库3的一部分的数据影印成为了可能。

这两个数据库和其它27个数据库组成了CMGPD-LN，这些数据的录入分四个独立阶段进行。在第一阶段，跨越了1980—1989十年，我们完成了CMGPD-LN数据库1的大部分和CMGPD-LN数据库3的一部分共100，000条记录的编码。在第二阶段，从1990到1999，我们获取并完成了CMGPD-LN数据库2、CMGPD-LN数据库3的剩余部分和CMGPD-LN数据库4至CMGPD-LN数据库10，总计新增400，000条记录的编码。在第三阶段，从1999年夏天开始，我们获得并完成了剩下的19个CMGPD-LN数据库，新增一百万条记录的编码，并且花了一整年的时间来清理和重新整理全部一百五十万条记录。在最终阶段，从1999年到2006年，与第三阶段的工作有些重叠，我们访问了大约57个辽宁村庄，其间我们获得了组成我们本地版数据的250个相关数据库[2]。

仅CMGPD-LN数据库数据的获取和录入工作就持续了25年时间，我们受惠于超过12名研究助理的帮助，才将所有数据编码为机读格式。我们特别感谢Alice Suen和Anna Chi陪伴了我们第一个十年[3]；Sue Mei Tsay在第二个十年里陪伴了我们很久；孙惠成、肖星和激扬在最近的十年中对CMGPD-LN数据库的大部分数据进行编码；陈蔚然，我们在辽宁农村田野工作的助理和北京的项目经理[4]；台湾中央研究院的赖惠敏资助了数据库2一部分的录入。

[1] 犹他家谱学会藏人口册的影印版本和辽宁省档案馆藏的户口册的分析目录，以及 CMGPD-LN 中包含的 29 个行政区域的拼音和汉字列表，请参考附件 A 和附件 B。

[2] 在 1999 到 2006 年期间，我们在 Gao Jing 以及辽宁省地方志办公室、当地地方志办公室同事们的陪伴下做了八次田野调查。期间我们访问了 57 个大型农业村落，每天派 250 个人去拜访 CMGPD-LN 中人口的后代，搜集当地资源如家谱、墓志铭、契据和其它有关这些人口的家庭文件。我们把这些数据同 Campbell 和 Lee 2002a、 Ding, Guo, Lee, 和 Campbell 2004 的户口册数据做了一个初步比较，打算在未来地方数据的发布时做一个更详细的目录和描述。相似甚至更加本土化的工作可以参考 Nie 1992 年对大连附近一支辽宁旗人群体的研究，这支旗人没有被记录在盛京内务府的户口册里。

[3] 见 Lee 和 Campbell (1997, xix)的鸣谢部分。

[4] 其他参加数据输入的人员包括编码了数据库 3 的 Chen Jianhua ,编码了数据库 1 的 Robert Eng,编码了部分数据库 1 的 He Ti,编码了部分数据库 2 的 Alice Hsu,编码了数据库 2 的 Li Yan 和编码了部分数据库 3 的 Leng Yisheng,编码了部分数据库 1 的 Liu Guiping,编码了数据库 4 和数据库 7 的 Chi Shaolin, 编码了部分数据库 1 的 Julie Sun,编码了数据库 5 和数据库 10 的 Wang Xiaohong,编码了数据库 8 的 Yuping,

我们同样受惠于个人资助、校内基金和校外津贴，总计50万美元，这些是完成数据录入所必需的。资助来自：加州理工大学（the California Institute of Technology），洛杉矶加州大学（the University of California, Los Angeles），蒋经国国际学术交流基金会, the National Endowment of the Humanities, the National Program for Advanced Study and Research in the Peoples Republic of China, 密歇根大学人口研究中心和中国研究中心, the Wang Institute of Graduate Studies, 尤其是the Jeannette Lee Memorial Fund，资助了一半以上录入数据的费用，包括促进我们阶段三工作的所用费用。Barbara Calli、Susan Davis和加州理工大学的 Sabrina Boschetti，以及 Ruth Danner和密歇根大学的其他人员管理了这些资金，我们在此对他们的帮助表示感谢。

康文林先生负责了大部分程序的编写和计算机数据的准备工作，以及过去10年来所有的Stata程序，这些程序帮助编码员将原始数据录入到现在的CMGPD-LN中。在 1985-1987年期间，Lawrence Anthony写了一系列C语言程序，对数据库1死亡率和生育率做了初步估计。在1978年，康文林先生写了dBase III+程序和后来的dBase IV程序，使得原始数据同个人观测值相连，个人与他们的亲属相连，生成了CMGPD-LN数据库变量的雏形。在90年代初期，Chris Myers维护了这些程序并进行了拓展。然而，在90年代后期，为了将原始数据从Excel表格中输入、分析和生成新的分析变量，康文林先生用Stata写了新程序，因此我们废弃了原来的dBase IV programs程序。

2009年5月，当李中清先生任职于香港科技大学后，Susan Leonard 热心地接手了NICHD R01 HD057175-01A1 "Multi-Generation Family and Life History Panel Dataset" 并且促使数据通过ICPSR发布。

几位同事和学生花时间阅读并评论了使用指南。我们特别想感谢Susan Leonard、Matthew Noellert和汪琳岚。我们也想感谢我们5本书的合作作者和5本书的合作编辑，以及至今为止发表的使用CMGPD-LN的40多篇文章和章节的作者，包括Lawrence Anthony、定宜庄、Robert Eng、郭松义、赖惠敏、Chris Myers、Tan Guofu、王丰和我们在the Eurasia Project in Population and Family History的同事,特别是Tommy Bengtsson，以及George Alter、Marco Beschi、Renzo Derosas、Martin Dribe、Akira Hayami、Satomi Kurosu、Christer Lundh、Matteo Manfredini、Michel Oris、Noriko Tsuya。我们还要感谢与我们一起合作《Life Under Pressure: Mortality and Living Standards in Europe and Asia, 1700-1900》和《Prudence and Pressure: Reproduction and Human Agency in Europe and Asia, 1700-1900》的王丰[5]。在过去15年中，我们在一起进行了大量的研究，共同享受着合作与丰收。

CMGPD-LN的准备工作和通过ICPSR DSDR的公共数据发行是由the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)资助的，立项编号为R01 HD057175-01A1 "Multi-Generation Family and Life History Panel Dataset" from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), 资金来自the American Recovery and Reinvestment Act。数据库的内容仅仅代表作者个人意见，责任自负，不代表NICHD的官方意见。

和编码了部分数据库 1 的 Wang Yuanqing。

[5] 使用 CMGPD-LN 的完整的学术著作列表请见 ICPSR CMDPG-LN 网站。

本使用指南中文版最初由上海交通大学刘怡起稿，董浩、蒋依伶、李春圆、马馨、陈熙、宁雯雯、黄匡时、范长煜、彭杰、苗瑞凤对初稿进行了第一次修订。后由臧晓露完成对使用指南中文版的补充和更新，陈爽对成稿进行了最后修订。

对以上个人和机构，我们表示由衷的感谢。

引言

中国多世代人口数据库——辽宁部分（CMGPD-LN），包含了 150 多万条记录，提供了 1749-1909 年间生活在辽宁省的超过 260，000 位当地居民的社会经济状况、人口及其它相关信息。这些居民隶属于表 1 中列出的 29 个行政单位，包含了分布于辽宁北部、中部、中南部、辽东南部等现今位于辽宁省核心走廊区域的约 700 个村庄[6]。至今，我们已经将其中的 280 个村庄定位（地图 1）。

本指南首先对形成这一人口数据库的制度背景、数据库中的基本变量以及数据库中特定的旗人群体进行分析性总结，同时强调这一资料对于历史学和比较社会学的重要意义。本说明分为五个部分：第一部分简要介绍辽东地区八旗户口册的情况以及形成这一资料的制度背景；在第二部分，我们总结前人对于辽东八旗户口册进行的学术研究分析，以及这些研究成果对比较人口学和家庭史、清朝中后期历史、比较历史社会学以及社会分层比较研究的重要意义。第三部分中，我们对所有变量进行了详细讨论；第四部分，我们总结了 CMGPD-LN 中三种主要的人群（正身旗人、特殊义务旗人、皇庄官庄所属旗人）所对应的不同制度背景；最后，在第五部分中，我们对构成 CMGPD-LN 数据库的 29 个分支人口数据库进行逐一描述，以便使用者了解它们各自的特点和背景。

表 1 CMGPD-LN 包含的 29 个数据库

DATASET	户口册名称**	第一本可用的户口册	最后一本可用的户口册	卷数	观察值数量	个人数量
1	道义屯户口册-汉军人丁户口册	1759	1909	35	120,747	18,884
2	盖州户口册-汉军人丁户口册	1762	1906	27	42,819	6,957
3	打蜜户口册-汉军人丁户口册	1759	1909	32	32,106	5,013
4	城内户口册-汉军人丁户口册	1789	1909	27	55,678	10,054
5	盖州棉丁户口册-汉军棉丁户口册	1789	1909	25	56,048	10,714
6	牛庄刘二堡喇嘛园人丁户口册-汉军人丁户口册-汉军鱼丁户口册	1777	1906	25	54,870	10,208
7	肥地一面城人丁户口册-汉军人丁户口册	1756	1909	39	71,365	9,543

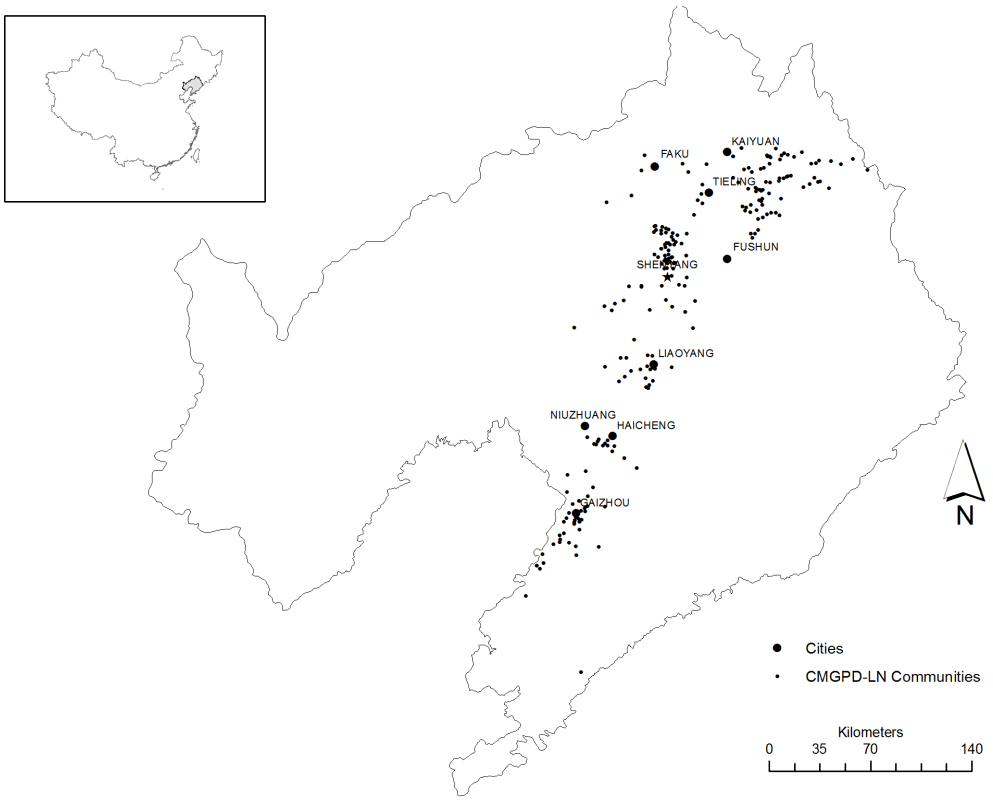
[6] 辽东地区主要指现今辽宁省境内辽河下游东部地区。这一地区有肥沃的辽河冲积平原，在十七世纪时一直作为未入关前清帝国的核心地带。然而辽宁省从行政上又分为辽东和辽西——即辽河以西的地区，清朝皇家内务府负责对这两个地区进行独立的管理；其中辽西地区主要受北京的皇家内务府管辖，辽东地区则受位于盛京（现沈阳）的内务府管辖，因此 CMGPD-LN 中所囊括的人群来自辽东地区。

8	盖州人丁户口册-满汉人丁户口册	1753	1909	21	53,284	11,952
9	大甸子户口册-汉军人丁户口册	1756	1909	27	76,986	14,060
10	郭三屯人丁户口册-汉军人丁户口册	1774	1909	34	35,081	4,932
11	八棵树老古洞人丁户口册-满汉人丁户口册	1759	1909	32	48,760	8,247
12	大辛屯人丁户口册-汉军人丁户口册	1749	1909	29	86,935	15,748
13	年马岱海寨人丁户口册-汉军人丁户口册	1749	1909	31	58,224	10,746
14	长寨子户口册 - 汉军人丁户口册	1768	1909	27	48,800	9,400
15	照化屯人丁户口册-汉军人丁户口册	1774	1909	26	50,862	8,758
16	貂皮屯人丁户口册 - 汉军人丁户口册	1768	1909	30	80,512	14,046
17	郎家堡魏家屯汤岗子海州岳州人丁户口册 -- 汉军人丁户口册	1756	1909	25	47,331	9,092
18	王指挥屯人丁户口册 -- 汉军人丁户口册	1765	1909	28	60,334	10,054
19	阿尔吉善户口册	1813	1909	20	15,098	2,517
20	海州户口册 - 海城汉军人丁户口册	1759	1909	26	119,186	20,032
21	王多罗树满洲牲丁户口册	1864	1906	9	9,494	2,300
22	王多罗树满洲人丁户口册	1792	1909	20	23,233	4,899
23	瓦子峪人丁户口册--汉军人丁户口册	1777	1909	27	75,042	12,884
24	五户人丁户口册 -- 汉军人丁户口册	1789	1906	27	44,790	7,687
25	棉花盐靛人丁户口册—佐领下盐靛庄头壮丁户口册	1756	1909	29	76,738	13,404

26	苏拜公人丁户口册	1864	1909	10	4,162	879
27	开档户口册	1810	1852	12	8,835	2,717
28	开档投从抱养人丁户口册	1792	1888	25	30,137	5,087
29	棉花盐靛小门人丁户口册	1861	1909	10	25,900	5,277
			总计:	732	1,513,357	266,091

* 虽然中国多世代人口数据库——辽宁部分（CMGPD-LN）包含了 29 个行政单位的数据库，第 29 个行政单位的人口实际是第 25 个行政单位人口的子集，因此第 29 个行政单位的人口可以与第 25 个行政单位的人口连接起来。

**各个行政单位户口册的名称是从户口册封面直接获取的。1864 年之前，各个户口册的名称往往以地名加“人丁户口册”的形式出现。从 1864 年开始，八旗系统把人口分类也加进了封面页，各个区域的户口册名称因此变成了地名加“汉军人丁户口册”、“满族人丁户口册”或者“满汉人丁户口册”。除了名称上的变化，各个行政单位的人口不变。在这张表中，我们列举了所有行政区域的户口册的名称。



注： FAKU: 法库 KAIYUAN: 开原 TIELING: 铁岭 FUSHUN: 抚顺 SHENYANG: 沈阳 LIAOYANG: 辽阳 NIUZHUANG: 牛庄 HAICHENG: 海城 GAIZHOU: 盖州 Cities: 城市 CMGPD-LN Communities：CMGPD-LN 村落 Kilometers：千米

地图 1 CMGPD-LN 中已经被定位的村落

1 辽东八旗户口册

CMGPD-LN 数据库所载的人口记录来源于保存在辽宁省档案馆的八旗户口册，世界各地都能通过美国犹他州家谱学协会[7]获得其中的部分电子数据。如表 1 所示，这些人口信息来源于 732 册户口册[8]。

CMGPD-LN 中的人口属于八旗下的盛京内务府。八旗是清朝满族统治者建立的一种兵农合一的系统，集行政、军事、经济功能于一体。这一制度由努尔哈赤及其子皇太极于 17 世纪早期创建并完善，逐步取代满族的部落组织，成为了清朝的前身满洲国家制度得以建立的立国之本。。八旗组织由不同的旗色区分为正黄旗、正红旗、正白旗、正蓝旗、镶黄旗、镶红旗、镶白旗、镶蓝旗。到清朝统治者于 1644 年入关前，辽东地区居住的所有人口均编入八旗之下。针对不同的民族，分设满洲、蒙古、汉军八旗，分别管辖满、蒙古、和汉军以及其他一些少数民族[9]。旗人出则为兵，入则为农，生活、生产、征战均在八旗组织的控制之下。,，清统治者入关后，原来居住在辽东地区的旗人也大量入关, 进驻北京城。尽管清廷在全国重要城市和战略要塞建立八旗驻防，八旗人口主要聚居于三个地区：北京城[10]；今宁夏、内蒙古、外蒙古等地；以及在中国东北辽宁、吉林、黑龙江等地区[11]。

CMGPD-LN 中的数据主要来源于居住于东北地区的旗人，他们由盛京内务府统领[12]。这部分八旗人口除少数是在清朝建立前迁移到辽东外，大部分是在 17 世纪晚期到 18 世纪初从山西、河北、山东等地移民到辽东地区的汉族人的后代(Ding, Guo, Lee, and Campbell 2004)。当他们在辽东定居以后，清政府把这些移民编入了八旗系统下的盛京内务府，使其成为内务府的世袭劳动力。

我们的数据库是从八旗盛京内务府编制的户口册录入的。在这一制度背景下，这些户口

[7] 关于 Genealogical Society of Utah 的介绍及如何在世界范围内获得进入数据库的许可，详情见 http://www.gensocietyofutah.org/ 我们要特别感谢 Melvin Thatcher 和他的同事们在过去二十五年间对我们收集、整理、分析数据提供了极大的便利。

[8] 现今留存在盛京内务府档案馆内的户口册大约有 4000 户，其中一半左右保存完好，其余则均有不同程度损毁。在完好的这部分户口册中，大部分户口册的内容都可以从 Genealogical Society of Utah 及其众多分支图书馆获得。我们估计仍有约 10%保存完好的户口册尚未被辽宁省档案馆影印记录，还需要等待它的最终完成。另外，许多“不完整”的户口册被题名，并同保存完好的户口册关联着，我们衷心地希望这些资料可以尽早被影印以便利于家谱学和历史学的研究。

[9] 除了这三种主要的民族以外，CMGPD-LN 中还有朝鲜族、朝鲜满人、高丽人或满洲高丽。

[10] 研究北京城八旗制度及整个中国大陆范围内的戍卫八旗系统请参阅 Crossley 1990; Ding 2003; and Elliott 2001

[11] 分布在北部和东北部的旗人人口组成了绝大多数的旗人，但遗憾的是目前对于这一人群的研究远少于对那些戍卫旗人的研究。详细请参考 Lee and Campbell 1997 and Ding, Guo, Lee，Campbell 2004 中对于东北部旗人的初步研究也包括在本手册中。

[12] 内务府是清代建立的用来管理皇家园林和其相关设施的机构。当时主要存在两个内务府：北京内务府和盛京（位于沈阳）内务府。皇家内务府不仅管理园林内的各种设施，在外还经营诸如盐业、人参、宝石翡翠等行业垄断以及为皇家收集海内外的珍宝(Bartlett 1991;Crossley 1999 and 2002; Qi 1998; Rawski 2001; Torbert 1977)。

册记录了八旗、盛京内务府、以及辽东地区--中国东北唯一在柳条边[13]以内的区域—的特点。人口登记及分类工作是晚期帝国社会的一个重要特征，特别是在中国东北这样一个国家政权个人在地方社会的建置过程中起了不同寻常的主导作用的地区(Chen 2009)。因此，这些数据详细记录了这一人群的人口、经济状况、社会生活等信息，其细致程度远胜于其它记载封建社会晚期的资料[14]。

1.A 盛京内务府

清政府远在征服中原之前便设立了盛京内务府来管理沈阳的皇宫及其附属设施。特别是位于辽东的皇庄以及皇陵[15]。盛京内务府下属大量的服务人员，包括工匠、农民、官员、士兵。每种人群对于皇家负有不同的义务，也具有不同的晋升机会。这些人组成了旗人社会从下至上的不同阶层，从旗的首领—佐领，到普通的旗人，即组成 CMGPD-LN 中的主要人群——正身旗人，再到向皇家上交棉花、鱼、猎物、蜂蜜、盐等专门负有特殊义务的旗人，以及被称为“小门”的底层的人、被称为仆人的雇工、社会最底层的奴仆。所有这些人群都归八旗中的正黄旗、镶黄旗、正白旗的三旗统领管辖，其他五旗则不归盛京内务府管辖，自然也没有出现在 CMGPD-LN 中。

在盛京内务府管辖的一个半世纪中，辽东旗人约有一至两百万之多[16]，大约相当于当时全省人口的十分之一[17]，辽东旗人数目的四分之一。在 1741 年至 1897 年间，辽宁省登记的人口以年增长率近百分之二的速度，从少于四十万增长到接近五百万。如此高的增长率是持续的移民带来的结果。在清朝统治的一百多年间，从河北、山西、特别是山东省等地不断有人口迁移到辽东[18]。现今大部分的辽宁本地人认为其祖先是在十七世纪中期到晚期移民到辽东地区的，实际上，许多人是在更晚的时候才迁移进来的[19]。当然，在这些移民到来之前，

[13] “柳条边”是清政府在 1681 年划立的用来防止中原地区的汉族人口迁移到中国东北地区而侵犯旗地的边界。“柳条边”是由两座平行的，相距 3.5 米，高 1 米的泥墙和之间的柳枝篱笆组成(Edmonds 1979)，如地图 1.1 中所示，“柳条边”主要分成东西两部分：西面部分从长城脚下蔓延 700 公里到松花江河畔；东面则从开源北部一直向东南延伸到朝鲜边境。尽管清政府一再宣布禁止民众私自迁移跨越“柳条边”，但整个清代年间，一直存在汉人暗中跨越柳条边而进入吉林和黑龙江等省份(Lee 1970; Reardon-Anderson 2005)。清政府对于这些人并没有驱逐，反倒是对其进行登记从而使这些人的定居合法化。“柳条边”以北的地区在制度上有很大的区别，清代记录中也称这些人为“边外”。

[14] 参阅（Lee and Campbell 1997）对数据库中的“道义屯户口册汉军人丁户口册”这一辽东地区人群所做的最初分析。对其他地区人群进行的相关研究细节亦可参阅本说明中第二部分中对相关研究的总结（Ding, Guo, Lee, and Campbell 2004）。

[15] 对皇家陵墓的研究较多，而对其下的贵族陵墓的研究则比较少。相关研究可参阅（Feng Qili 1996）对主要位于北京附近，超过 240 个清代王公贵族陵墓进行的细致调查。

[16] CMGPD-LN中所收录的732卷户口册所涉及的人群占所有现存户口册人口总量的五分之一，而这部分人群是皇家内务府所记录的所有户口册中，平均每个个人的登记与事件观察数登记最密集的部分，所以推测在剩下部分的户口册中涉及的人口最少有一百万甚至更多。

[17] 大部分人口居住在国家经营的农庄中，即官地，主要受盛京赋税部门管辖。

[18] 绝大多数的家谱是从生活在辽东半岛的家庭中收集来的，这些人称自己最初是从山东半岛东北部的登州黄县一代迁来的。

[19] Ding and Liu 1988详细介绍了后期的移民的情况。

本地确实也有一些原住民[20]。清政府把十七世纪辽东地区的大部分人口编入了八旗系统，同时把汉族移民纳为汉军[21]。尽管，这些人实际上是汉族，但他们强调自己区别于后来定居在此地而不属于八旗系统的汉人。从这些人后裔的口述资料可以看出，尽管已经进入了二十一世纪，这些 CMGPD-LN 中记载的人口的后裔依旧认为他们与生活在辽东的其他汉人不同，称自己是旗人，而称他们为民人或平民。另外，入旗的过程也被称作“随龙”，字面上理解即跟随龙的意思[22]。

八旗组织大概的制度背景已清楚[23]，绝大多数受盛京内务府所管辖的旗人都是世袭的农民，他们出卖劳动力或者是上缴固定的租金而换来土地的使用权及其他一些特权。大多数的农民向盛京内务府上缴粮赋，有的则上缴各种各样的农产品或是履行特殊义务，典型的如：上缴棉花、染料、鱼、水果、蜂蜜、盐以及或服不同的特殊义务、担当守卫等。尽管如此，这些义务同旗人身份带来的益处相比则显得微不足道。分配给他们的土地通常是可观的，是全国普通水平的几倍到多倍[24]。

更多的是，受盛京内务府所辖的这些农民还有机会享受到八旗系统中对于个人晋升、财政安全所带来的优势[25]。作为一个旗人，他可以从事多种职业，从工匠、士兵、平民到军队的长官等等[26]。这些职务带给他们丰厚的薪酬、各种各样的权利及特权，甚至能有差别地被纳入政府为八旗社会提供的社会安全保障系统中，该系统一方面保护孤儿及寡妇，另一方面也为老人、退休和失业人员提供帮助。

---

[20] 这部分人存在但数量有限，Cao Shuji 1997 曾估计 300 万中国东北居民中有 250 万在 17 世纪初明清过渡期间迁出，留下的人群又有很多在清朝征服过程中丧生。剩下的很大一部分人最终都被归入了八旗系统，他们和 CMGPD-LN 中人群享有类似的制度背景。

[21] 这其中有八处例外：在21、22号数据库中的王多罗树人群全部是满族，11号数据库中的八棵树人群也包括了一定数目的满族群众；数据库8中的盖州满汉和25、29号棉花盐靛庄头壮丁则包括了一些被归为朝鲜族的个人；26号苏拜公人丁则全部是蒙古族。此外，数据库16中貂皮屯人丁虽被划为汉军人丁，其中的几户也被认为是满族。

[22] 可参阅Nie 1992对八旗中非CMGPD-LN所涉人群，在对其他五旗的一些后代进行的研究中所得到的类似结论。

[23] 详情可见 Crossley 1999, Ding 2003, Elliot 2001, 尤其 Lee and Campbell 1997 和 Ding, Guo, Lee, Campbell 2004 等研究辽东人口的资料；也可参考 Chen 2009 对位于哈尔滨城南双城县内的 120 个边外村庄在特定历史时期和制度背景下的情况所做的深入研究。

[24] 十八世纪，当时全国人均拥有土地面积少于 6 亩（相当于一英亩），而辽东地区作为旗人的农民却拥有 30 亩土地（相当于 5 英亩）Ding 2003。在关外北部，黑龙江双城的旗人人口人均土地有 60 到 120 亩之多，是全国人均水平的 10 到 20 倍以上，并且拥有土地的多少也和其的身份相关， 比如来自于辽宁农村的旗人和来自于北京这样的大城镇的旗人拥有的土地面积则是不同的（Chen 2009）。

[25] 参考 Lee and Campbell 1997 中对这些优惠政策进行的细致分析和其在人口领域产生的影响。

[26] 特别是在盛京内务府内部更容易找到职位，根据一个在十八世纪晚期的表格看，大约有超过 900 名被雇佣旗人直接受总经理（总管大臣）进行管辖：包括 3 名 Booi 旗人，3 名骁骑校, 1 名内管领, 1 名长领达， 1 名堂主事, 1 名委主事, 15 名笔帖式, 5 名司库或副司库, 16 名库使, 5 名食俸催长, 3 名项带领催, 和 815 名执事人、甲丁和牧丁。可参阅 Tong Guan （1994.23）引用的 Heitudang vol 731。CMGPD-LN 中的一些旗人还通过最普遍的方式—科举，从而加官晋爵而走上高位。

旗人的身份同样也使他们的活动受到了一些限制。旗人和非旗人虽然住在一个融合的居民社会中，但却被规定所分隔开，旗人不能与非旗人之间通婚，也不能售卖、租借其旗人权利给一个非旗人。同时，对于这一研究的目的更重要的是，旗人首先属于他们的行政人口单位，第二才属于他们所登记的地方，即现今的单位与户籍。他们只有通过正式的申请通过批准后才能更改地址。虽然这一情况在清朝时期全国各地普遍存在，但这些规则在旗人社会中异常严格。未经授权的迁移会在八旗和内务的户口册中登记为“逃亡” (Campbell and Lee 2001)。我们在 CMGPD-LN 中很难找到一个改变行政隶属的个例。

CMGPD-LN 所记录的数据是特别完整和精确的，这是因为其所记录的圈子是一个在多方面相对隔离的小社会。它涵盖的时间跨度久达一个半世纪，从 1749 到 1909 年，提供了一个七代人或许更多代的多世代人口资料。通过这一平台，我们能够观察这一时间历程中的个人、家庭、居民、血亲、当地社会组织、以及地方化水平，使我们可以把这超过 26 万人定位在 50000 户、10000 个户组、1000 条血亲网络、以及超过 700 个村庄中，来观察与比较个人与群体的行为，更好地从整体上实现对中国历史和社会科学的了解。

1.B 盛京内务府与八旗人口登记

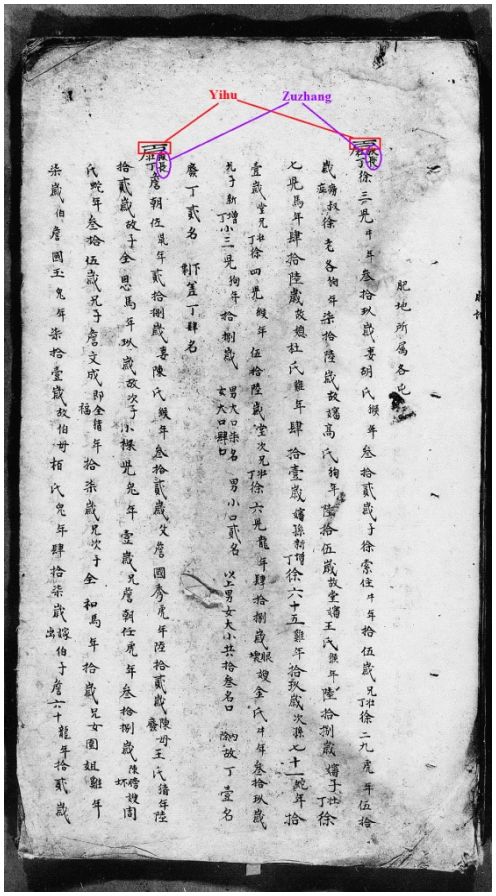
八旗人口登记在八旗制度建立的初期就存在，到 18 世纪早期逐渐正式化。1727 年，雍正皇帝规定人口登记应该以户为单位来管理。因此，人口登记中不应只记录每一个旗人成员的身份地位信息，还应包括这一户中所有其他的男性亲属。另外，每过三年，八旗管理人口的机构需要更新这些记录，并向中央政府上交一份复制本。在 1730 年代间，乾隆皇帝进一步规定将旗人人口按其类别记录，因此产生了一种等级制度的人口登记系统，这些规则对旗人之间的政治经济特权进行了区分。最后，乾隆于 1748 年在盛京内务府下面设立了一个中央机构来专门收集与保管旗人的人口登记记录(Tong and Guan 1994)。

一般而言，在八旗制度中有两种形式的登记：人口登记和“三代”登记。人口登记是包含每一个人的完整清单，包括姓名、头衔、年龄、以及与人口相关的所有重要信息，以户为单位将个人组合起来。而“三代”登记则只是记载男性成员及其父亲和祖父的姓名与社会头衔，以个人与父系血统来组建清单。同人口登记相比而言，“三代”登记所包涵的人来自于更广泛的社会阶层，包涵的人数也更多——可能有七百万之多。但其所提供的每个个人的信息则太过简略。

CMGPD-LN 数据库中的信息全都来源于上面提到的这些人口登记。就数据库中信息所涵盖的时间段而言，记录主要以两种模式对人口进行分组：在 1789 年之前，人口登记以“族”为单位；从 1789 年起，登记则以居民的“另户”为单位[27]。图 1 中所显示的就是在 1789 年前的一页户口册，这页肥地一面城的人口信息登记于 1783 年。所显示的这页中，有两户人家，一户是完整的，另一户则是部分的。在每一户的记录顶端都写着族长的名字，这意味着这里的一户实际上是包括了一族人。图 2 中，则是在 1792 年对同样人群的人口登记，这份登记中虽然也记载了 1783 年户口册中同样的人，但按照一种新的模式，即包括了三个另户，并且每个另户的户主也记录在册。

[27] 参考Liu 1998 , 13-24 and 152-162 两篇文章对于旗人“户与户组”，主要是在十七世纪间的情况的研究。

尽管表中记录的基本单位从族变成了户，户口册总体上还是按照统一的标准所编排。记录表中的人群主要按照居住的村庄、族或户来分组。每三年，所有旗人的下面这些信息都要按顺序记载在表中：包括同族长或户主的关系、姓名、成年旗人的身份、年龄、生肖[28]、阴历生日、出生时刻、婚姻、死亡、或者移民状况（如果在两次人口普查之间出现）。如果是成年男子，还包括其是否残疾，其亲属族长的姓名、归属八旗中哪一旗以及居住的村庄。每一个个人按其与户主的关系排一列：户主的妻子、儿女、孙子女排在最前；接着是他的兄弟及他们的后代；然后是他的叔舅、姑姨、表兄弟等[29]。另外，户主的确立要根据其父亲、祖父的名字及地位，寡妇则依据其丈夫的姓名和地位。总而言之，记录甚为全面，那些未经允许离开当地的人或者是失踪的人都会被标记为逃丁。一般而言，在每一次的人口登记中，新出现的条目只可能是因为外人的迁入或者是新生命的降生。



[28] 生肖即中国人把12年的一个轮回中的每一年按照一种动物来区分：鼠牛虎兔龙蛇马羊猴鸡狗猪。

[29] 一般情形下，妻子的名字紧跟在其丈夫的后面。但有一种情况，即男方的父母亲，特别是老母亲还在家中的话，她的名字会跟在户主（她的儿子）的后面，儿媳的前面。

图 1　肥地一面城 1783 年户口册式样

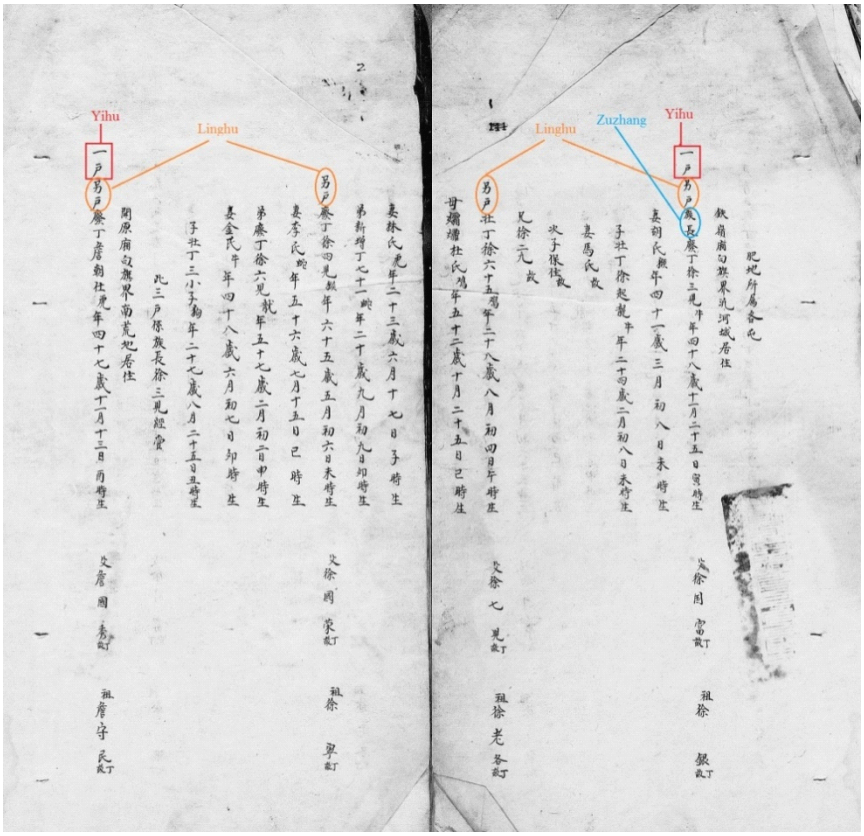


图 2 肥地一面城 1792 年户口册式样

1.C 从户口册到中国多世代人口数据库——辽宁部分（CMGPD-LN）

把原始户口册转化为中国多世代人口数据库——辽宁部分（CMGPD-LN）的第一步就是把户口册的信息由人口录入入到数据库文件中。从 20 世纪 80 年代录入入道义户口册开始，录入工作直到最近才完成。本指南“鸣谢”部分提供了录入工作的时间进度表，以及编码员和数据库的相关信息。从 20 世纪 90 年代早期开始，每一本户口册都被录入在一张独立的电子记录表中。例如，图 3 显示的是图 2 中的信息被录入后的式样。编码员把每个系列第一本户口册的内容直接输入到一张空白的电子记录表中，每条记录在 case 中按顺序编号。被登记为死亡、逃走、外嫁、过继出户的人口，编码员把他们的代码录入在 vital1 和 vital2 中。编码员把每个个人与户主的关系记录在字符串变量 rhhead 中。

	A	B	D	E	F	G	I	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	AB	AC
1	case	date	clan	hid	yihu	rhhead	addr	sex	gen	rank	occu	mari	name	vital1	vital2	birthyr	age	month	day	hour	disab	lcase	ldate
2	1	792	1	1	0	e	1	2	3	1	7	1	xu saner	0	0	2	48	11	25	3	0	1	786
3	2	792	1	1	0	w	1	1	3	0	0	1	hu shi	0	0	9	41	3	8	8	0	2	786
4	3	792	1	1	0	ls	1	2	4	0	1	3	xu qilong	0	0	2	24	2	8	8	0	3	786
5	4	792	1	1	0	lsw	1	1	4	0	0	0	ma shi	1	0	0	0	0	0	0	0	4	786
6	5	792	1	1	0	2s	1	2	4	0	0	0	baozhu	1	0	0	0	0	0	0	0	0	0
7	6	792	1	1	0	lob	1	2	3	0	0	0	xu erjiu	1	0	0	0	0	0	0	0	5	786
8	7	792	1	2	0	e	1	2	3	0	1	1	xu liushiwu	0	0	10	28	8	4	7	0	8	786
9	8	792	1	2	0	m	1	1	2	0	0	3	du shi	0	0	10	52	10	25	6	0	7	786
10	9	792	1	2	0	w	1	1	3	0	0	1	lin shi	7	0	3	23	6	17	1	0	0	0
11	10	792	1	2	0	lyb	1	2	3	0	3	2	qishiyi	0	0	6	20	9	9	4	0	9	786
12	11	792	1	3	0	e	1	2	3	0	7	1	xu sier	0	0	9	65	5	6	8	0	10	786
13	12	792	1	3	0	w	1	1	3	0	0	1	li shi	7	0	6	56	7	15	6	0	0	0
14	13	792	1	3	0	lyb	1	2	3	0	7	1	xu liuer	0	0	5	57	2	2	9	0	11	786
15	14	792	1	3	0	lybw	1	1	3	0	0	1	jin shi	0	0	2	48	6	7	4	0	12	786
16	15	792	1	3	0	lyb1s	1	2	4	0	1	2	sanxiaozi	0	0	11	27	8	25	2	0	13	786
17	16	792	2	1	0	e	2	2	3	0	7	1	zhan zhaoshi	0	0	3	47	11	13	10	0	18	786
18	17	792	2	1	0	f	2	2	2	0	0	0	zhan guoxiu	1	0	0	0	0	0	0	0	16	786
19	18	792	2	1	0	w	2	1	3	0	0	1	zhou shi	0	0	6	44	1	14	10	0	19	786
20	19	792	2	1	0	ls	2	2	4	0	1	0	zhan wencheng	1	0	0	0	0	0	0	0	20	786
21	20	792	2	1	0	lsw	2	1	4	0	0	1	zhao shi	0	0	1	25	7	17	1	0	21	786

图 3 图 2 中肥地一面城 1792 年户口册式样的数据录入

我们遵循的基本原则是：在录入过程中，编码员尽可能忠实地录入原始户口册的内容。其它过程，如数据的分类、重新编码、清理和创建分析变量尽量由计算机程序来完成。当原始户口册的内容很难被直接录入时，编码员遵循一系列明确的规定进行录入，做到无论是哪位编码员输入，都能不需要任何主观的阐释就能与原始户口册表达相同的意思。例如，根据一系列明确的规定，关系（relationship）从原始户口册中的汉字描述被转录为由字母组成的字符串。无论哪位编码员录入，只要按照这些规定，输入的内容与原始户口册的内容是一致的。代（generation）的编码是根据关系（relationship）变量有电脑程序生成。

当编码员把地址、职位和残疾信息转换为数值变量输入户口册记录表时，每当他们遇到一个新地址、职位或者残疾记录，他们都将其汉字和汉语拼音输入到另外一张编码表中，并按照次序顺次编码，然后在户口册记录表中输入这个新号码。因此，在一个新行政单位的户口册中遇到的第一个地址被标为 1，下一个遇到的新地址被标为 2，诸如此类。如果一个有了编码的地址再次出现，则输入它已经存在的号码。大体上，任何不能与编码表中已存在的地址、职位或者残疾记录完全一致的记录，都被赋予了新代码，即使它们唯一的区别明显是由于错别字或文字的缺失造成的，例如村名中少了一个字或用了同音字。由于录入过程中，编码员是以行政人口为单位逐一录入的，每一个行政人口数据库都独立编码，其数字编码均从 1 开始，因此，在不同的行政人口数据库中，相同的数字编码有可能代表的是不同的残疾类型、职位或者地址。

在录入一个行政人口的下一本户口册时，编码员打开一张新的电子记录表，复制他们刚刚完成的户口册，根据两本户口册之间相隔的年数更新年龄，复制 case 中的编号到 lcase，这样同一个人在新户口册里的记录就与其在旧户口册里的记录连接了起来。随后编码员们检查新户口册的电子记录表，根据新户口册的内容更新信息。他们为之前户口册没有记录过的人插入新条目，删除不再存在的人的条目，重新整理居住地变更的人的记录，更新新户口册表现出来的每个方面的细节。对新出现人口，编码员们通过给 vital1 赋予不同的值记录他们入册的方式。例如，一位女性 Zhang shi 由于嫁给本户男性 Sanxiaoer 为妻而入册，这位新入册妇女的 vital1 值就被设为 7。

	A	B	C	D	E	F	G	I	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	AB	AC
1	case	date	interpol	clan	hid	yihu	rhhead	address	sex	gen	rank	occu	marital	name	vital1	vital2	birthyr	age	month	day	hour	disabil	lcase	ldate
2	1	795	0	1	1	0	e	1	2	3	1	7	1	xu saner	0	0	2	51	11	23	3	0	1	792
3	2	795	0	1	1	0	w	1	1	3	0	0	1	hu shi	0	0	9	44	3	8	8	0	2	792
4	3	795	0	1	1	0	1s	1	2	4	0	1	3	xu qilong	0	0	2	27	2	8	8	0	3	792
5	4	795	0	1	2	0	e	1	2	3	0	1	1	xu liushiwu	0	0	10	31	8	4	7	0	7	792
6	5	795	0	1	2	0	m	1	1	2	0	0	3	du shi	0	0	10	55	10	25	6	0	8	792
7	6	795	0	1	2	0	w	1	1	3	0	0	1	lin shi	0	0	3	26	6	17	1	0	9	792
8	7	795	0	1	2	0	1yb	1	2	3	0	1	2	qishi	0	0	6	23	9	9	11	0	10	792
9	8	795	0	1	3	0	e	1	2	3	0	16	1	xu sier	0	0	9	68	5	6	8	0	11	792
10	9	795	0	1	3	0	w	1	1	3	0	0	1	li shi	0	0	6	57	7	15	6	0	12	792
11	10	795	0	1	3	0	1yb	1	2	3	0	7	0	xu liuer	1	0	0	0	0	0	0	0	13	792
12	11	795	0	1	3	0	1ybw	1	1	3	0	0	3	xu shi	0	0	2	51	6	7	4	0	14	792
13	12	795	0	1	3	0	1yb1s	1	2	4	0	1	1	sanxiaoer	0	0	11	30	8	25	2	0	15	792
14	13	795	0	1	3	0	1yb1sw	1	1	4	0	0	1	zhang shi	7	0	7	22	1	15	1	0	0	0
15	14	795	0	2	1	0	e	2	2	3	0	7	1	zhan zhaoshi	0	0	3	50	11	11	10	0	16	792
16	15	795	0	2	1	0	w	2	1	3	0	0	1	zhou shi	0	0	6	47	1	14	10	0	18	792
17	16	795	0	2	1	0	1sw	2	1	4	0	0	3	zhao shi	0	0	1	28	7	17	1	0	20	792
18	17	795	0	2	1	0	1s1s	2	2	5	0	0	2	quaner	0	0	11	6	3	13	6	0	21	792
19	18	795	0	2	1	0	1s1d	2	1	5	0	0	2	qiuju	0	0	9	8	9	1	10	0	22	792
20	19	795	0	2	1	0	2s	2	2	4	0	7	2	quanhe	0	0	7	21	7	17	10	34	23	792

图 4 肥地一面城 1795 年户口册前 2 页的数据录入

将相邻次序的两本户口册中的数据进行比较，可以看出所录入的数据的主要组织特点。图4显示了1795年肥地一面城户口册前两页的数据截图。除了图4记录的是三年后的信息以外，其出现的人与图2中1792年户口册和图3中相应电子记录表中出现的人大致相同。例如，户主 Xu Saner在图4中年长了3岁。其他人的年龄也同样基于户口册的内容更新。在1792年户口册中被登记为死亡的人，在1792年的户口册`vital1`中被设为1，在1795年的户口册中就不再出现了。Sanxiaoer的妻子Zhang Shi作为一条新记录被插入进来，她的`vital1`被设为7。`lcase`的值通过复制`1792`年`case`的值，这样就把其个人`1792`年的记录同`1795`年的记录连接了起来。

STATA程序将不同电子记录表的内容合并为一个数据库，自动将不同户口册中相同个人的记录连接起来，把个人与他们的亲属连接起来。由此我们依靠 `lcase` 变量把相同个人的所有记录连接起来，创建了PERSON_ID变量,用于识别不同户口册里相同的个人个人。我们基于`rhhead`，即CMGPD-LN中的RELATIONSHIP变量，还创造了FATHER_ID、MOTHER_ID、HUSBAND_ID 以及其它链接个人与其父母、其配偶的父母和丈夫的变量。根据RELATIONSHIP的值，连接程序识别了个人的父母、其配偶的父母或丈夫RELATIONSHIP的值，并将其归为同一户。例如，如果一个人是户主的长子(RELATIONSHIP `1s1s`)，程序就会找到户主的长子(`1s`)。如果成功找到了，程序会复制这个亲属的PERSON_ID到相应的位置。在连接户主长子与他父亲的例子中，程序会为户主长子的长子复制户主长子的 PERSON_ID作为FATHER_ID。

以上对各个特定亲属的连接只是第一步。接下来电脑程序将数据库中的人口进一步自动连接，并将他们归入各种层面的亲属关系网络。这些变量包括FOUNDER_ID, UNIQUE_GROUP和 UNIQUE_YI_HU。这些标记亲属关系网络的变量是为了将拥有同一个父系祖先的个人集合到一起。这些变量允许我们将不同定义的父系血亲组作为一个单位来分析。我们根据血亲组的不同定义，从狭义的到广义的，提供了几个不同的变量。最狭义的是根据FOUNDER_ID的值来定义的，紧接着是FOUNDER_INFERRED_ID，然后是UNIQUE_YI_HU，最后是UNIQUE_GROUP。UNIQUE_HH_ID 在构建时与上述变量有所不同。它将不同户口册中居住在同一户的人分为一组，因此它不能与本部分描述的其它血亲组变量直接相较。

在创建这些血亲组变量时，我们第一步通过上溯FATHER_ID的值重新构建男性的系谱。例如，我们首先复制个人的父亲的FATHER_ID来构建GRANDFATHER_ID，并以此类推。

这种用来识别父系祖先的操作一直重复到最早的已得户口册。户口册中最原始祖先的PERSON_ID变成了他所有子嗣的FOUNDER_ID。广义上的血亲组通过构建UNIQUE_YI_HU 和 UNIQUE_GROUP来定义，它们利用FOUNDER_ID识别个人最原始的祖先，若最原始的祖先在最早的已得户口册中生活在相同一户，这些个人将会被分为一组，或者对于UNIQUE_GROUP来讲，将拥有同一个姓氏的一户人连接起来。

与此同时，程序也将职业、残疾状况的原始信息和由编码员输入、用于分析的标识变量及其他变量的具体信息进行了诠释。这些过程大多是在拼音的基础上完成的。对于官方职位，编码员将原始信息的汉语拼音字符串中的辅助信息分离出来，只留下核心职位信息作为主要标识变量的基础。个别表示职位的标识变量基于手动分类。

2 使用 CMGPD-LN

2.A 概述

CMGPD-LN 数据库适用于人口统计学、家与家族、亲属关系、社会分层、卫生健康等研究领域中的多个课题。和以上领域中常用的通过当代固定样本跟踪调查跟踪中建立起来的公开利用的以个人为单位的追踪个人数据库一样，CMGPD-LN 追踪个人，在多个时点记录他们各方面的特点，并且提供了较为详细地这些个人及其家庭所经历的重要的经济、社会、以及人口学事件的时间。另外，和其他一些数据库一样，CMGPD-LN 还定期提供个人的家庭背景和社区信息。

CMGPD-LN 数据库最与众不同的特点是可以用于血亲网络与多世代人口发展过程的研究。CMGPD-LN 的独特在于它是目前唯一对外公开的、细化到个人层面的人口数据库。这是因为通过人工或自动化地链接户口册上的原始记录，数据库跟踪记录了许多家庭多达七代人的信息，并且重构了存在于家庭成员之外的父系血亲网络。因此，CMGPD-LN 数据库不仅能支持关于父母亲与子女、兄弟姐妹之间的特征、变化、以及结果等变量的相关度的分析，，还支持亲缘关系更远的亲属间的以上变量的相关度分析。

此外，CMGPD-LN 数据库为发展和评估目前一些先进的数量分析方法提供了一个极有价值的工具。因为每个个人都存在于多层面、多轴心的环境之中，从一户人家到整个地区，该数据可被用来衡量估计分层及其它模型的技术，这些模型考虑了观测在不同水平上聚类的效应。。除此之外，这些数据的时间跨度超过 150 年并且提供了个人的生命史，所以它还可以用于发展和评估年龄-时期-队列分析（age-period-cohort analysis）的方法的新技术方法。数据库中有大量的处于各种不同条件下的队列、时代、和年龄组，这些条件有助于区分影响。

下面，我们将介绍 CMGPD-LN 数据库的各项特色，同时也指明其存在的局限性，最后将提供一些应用该数据库对人口统计学、健康、及分层等领域中的重要课题进行研究的实例。CMGPD-LN 数据库的特点在于其具有前瞻性、截面性、纵向性、封闭性、详尽性、多层次、多世代的特点。同中国人口统计学领域在 20 世纪之前几乎所有其他资料不同，CMGPD-LN 记录了妻子或是寡妇的信息。当然，该数据库最重要的不足在于户口册遗漏了那些在婴儿期或是幼年时夭折的男孩及绝大部分女孩的信息。此外，表中对于社会经济状况的衡量也仅限于具有官方职位或头衔的男性。

2.B 特点

2.B.i前瞻性

CMGPD-LN 数据库中的记录具有前瞻性。每一册户口册描述了它编造时的情况。某种意义上讲，某一特定年份的户口册的数据相当于当代的追踪调查或者追踪人口统计在一波中搜集的数据。因为其反映的是当时的实际情况，所以就把选择性偏见或是记忆错误等因素所

造成的影响减到了最小。与其形成对比的是，那些长期以来作为研究中国古代血亲关系及人口的基本资料的家谱或宗谱则是一种追溯的记录形式。绝大部分家谱都只是定期地进行更新修订，并且依赖于家族成员对于多年甚至几十年前发生的人和事记忆。因此，家谱往往有选择性地记录其家族成员。那些已婚并且已有后代，或者是有成就的宗系成员最可能被后人记住并且载入家谱(Campbell and Lee 2002a)。

2.B.ii 纵向性

CMGPD-LN 数据库中所记载的数据具有纵向性的特点。数据库记录了个人一生的历程，在不同的时间节点记录下个人的各种特征以及其生命史中重要事件和转变发生的时间。就他们本身而言，每一册户口册并不具有纵向性的特征，其格式和组织更像每三年一次的人口普查。是编码员把某一个人在一个户口册中的记录与其在先前户口册中的链接起来的。幸运的是，这些户口册都以几乎相同的顺序连续记录个人，为编码员将相邻户口册里的记录链接提供了方便。我们还设计了程序来集合相邻户口册中成对记录的连接，并由此创建了标识符来标记不同户口册中对应相同个人的记录。不同的户口册中同一个人的记录自动生成的连接，是变量“个人编码”（PERSON ID）的基础，通过 PERSON ID 来区分个人在所有户口册中的相关记录。举个例子，对 PERSON ID 变量和 YEAR 变量进行排序，数据库即对所有数据先以个人，再以这一个人的不同年份的数据进行排序。同时 PERSON__ID 也是数据库中链接个人之间的信息的基础，比如 WIFE_ID 妻子编号, HUSBAND_ID 丈夫编号, FATHER _ID 父亲编号等变量。

2.B.iii 封闭性

CMGPD-LN 数据库中所涵盖的人群很大程度上是封闭的，简单的说就是这个群体中的男性、已婚女性以及丧偶女性很少有迁出群体及其他退出的现象。每一个成员离开的事件都在原记录中被标注，使得我们可以用一些标识性变量，诸如 NEXT_ABSCONDED（即将逃亡）、NEXT_DIE（即将死亡）、NEXT_MARRY（即将成婚）、MEXT_REMARRY（即将再婚）来标示将会发生在当前记录和下一次记录中间的离开事件。另外如果家庭在辽东区域内迁居，他们会报告其新的位置，所以仍旧保留在这个户口登记系统中。居住地的变化都可以通过比较同一个个人在相邻记录中 UNIQUE_VILLAGE_ID 变量的情况来获得。只有一小部分人彻底离开辽东地区。当这些人迁出时，户口册上都会加以注释这一事件，这样我们就可以将这些观察量截尾。NEXT_ABSCONDED 变量表明了某一个人在下一册户口册中会被注释为“逃”。当然，有些个人从数据库中看起来似乎消失了，即他们出现在某一年的户口册中却在我们能找到的下一册户口册中没有出现。事实上，大部分情况下这是因为记录他们离开时间的的户口册在保存的过程中遗失了。更重要的是，这些人的生命史基本都可以在他们最后一次的记录处截尾。封闭的人口不仅意味着我们可以追踪所有的男性、已婚女性及丧偶女性的整个生命史直至他们呢去世或者是最后一册户口册，还意味着我们可以追踪某一家庭中多代人的历史。

2.B.iv 详尽性

CMGPD_LN 涵盖的信息是极为细致的。它提供了个人的诸多信息，如人口、社会、家庭层面的重要特点、重要事件或者转变发生的时间等等。每一次人口登记记录了距离上次登记的三年间发生的死亡、出嫁及非法的潜逃事件。户口册还每隔三年记录了详细的家庭成员

与户主的关系，使我们可以容易地推断出这其间发生的婚嫁、生育等事件。对于成年男性，户口册还记录了其行政职位和地位。因此，个人获取官职、退休的时间都可以通过比较相邻户口册中地位头衔的变化来推断。

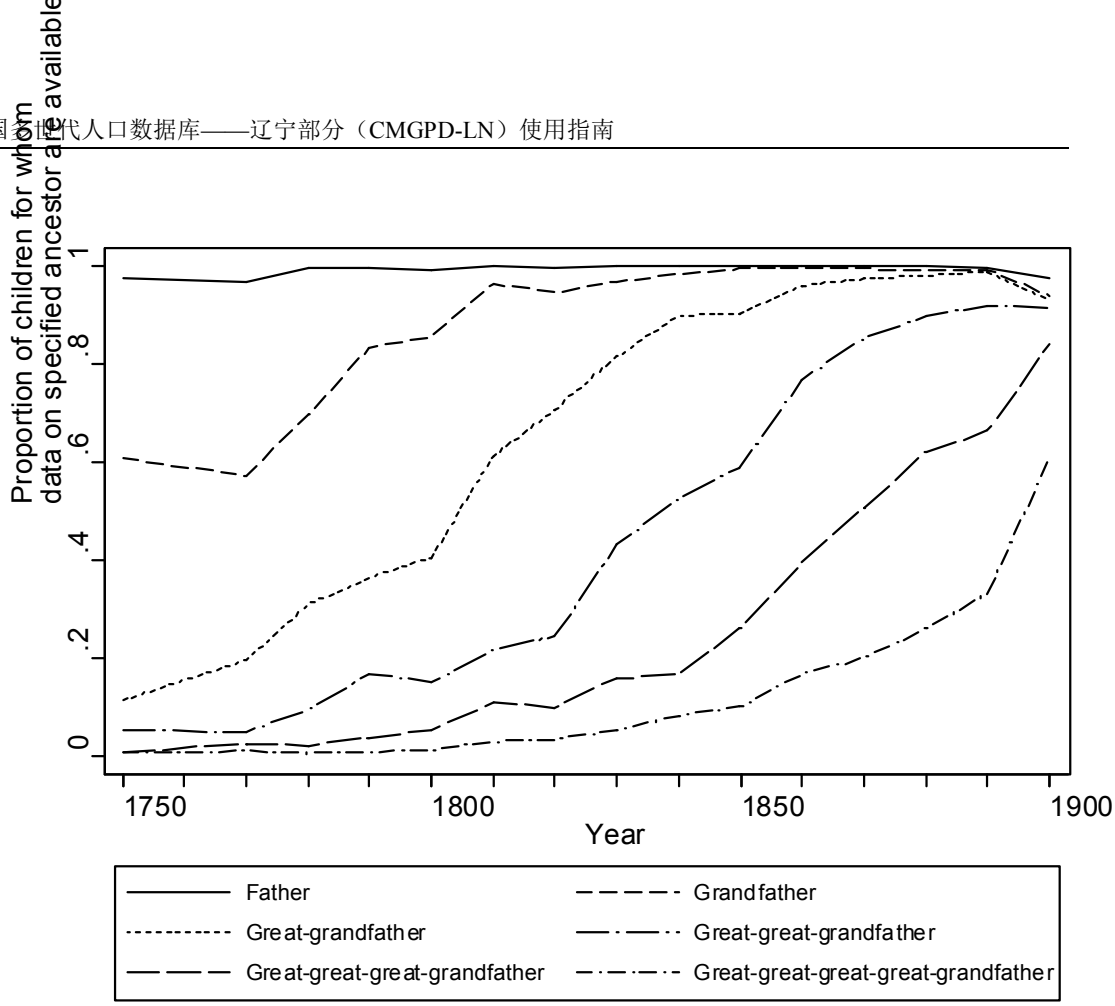
2.B.v 多层次性

CMGPD_LN 中，所有个人都置身于多中心或多层次交叉的环境中：共居的户、共居团体、父系血亲家族、社区、和地理区域。有些时候，这些环境相互重叠。例如，父系血亲通常会分布到很多个社区，每个社区中一般也有多个家族团体。CMGPD_LN 中的数据不仅可以应用来研究社会组织中不同层面的那些可以衡量的性质如何来影响某一个人在人口学及社会领域的行为结果，同样也适用于运用分层模型以及其他一些先进的方法在不同层次上来衡量或者考虑那些原本无法测量的变化。在发行的基本版数据中就包含了一系列标识符来对所观察数据按照层级模型来进行分组，帮助构建变量来衡量某一群体的总体特征。例如，通过 HOUSEHOLD_ID 变量可以找出和某一家庭相关的在某一年内发生的事件，UNIQUE_VILLAGE_ID 变量标识与八旗户口册中某一村庄相关的所有观察量，DATASET 变量标识某一行政人口下的所有观察量，DISTRICT 标识该村庄地址在当代所属的行政区划，REGION 则标识某个村庄在整个辽东区域的所在位置。发行的分析版数据中，还有进行血亲团体划分的标识符，可以把所有观察量按照父系血亲进行归类。

2.B.vi 多世代

用 CMGPD_LN 数据库可以研究多世代过程及远亲网络。通过电脑对记录的自动链接，子女和其父母链接起来，这些联系加在一起又重建了血缘继承关系及血亲网络。这些两代人间的链接是基本版中 FATHER_ID、MOTHER_ID、GRANDFATHER_ID 等变量和分析版中 UNCLE_COUNT 、 AUNT_COUNT 、 BROTHER_COUNT 、 SISTER_COUNT 、 MALE_COUSIN_COUNT、FEMALE_COUSIN_COUNT 等亲属连接变量以及许多亲属或血系标识符的基础。

图 5 中总结了 CMGPD-LN 数据库男性中可以通过记录链接追溯到其特定的父系祖先的比例。其基本步骤是链接原始数据链接父亲和儿子，从而上溯到某一个人的祖父、曾祖父及更早的祖先。在最早的户口册中，每一个孩子的父亲都已经有记录；从十九世纪开始，户口册中几乎所有的个人都可以找到其祖父的记录；及至二十世纪，80%以上的孩子可以追溯到五代之前的先人，有超过一半的可以追溯到六代之前。



注：Year：年份 Father：父亲 Grandfather：祖父 Great-grandfather：曾祖父 Great-great-grandfather：高祖父 Great-great-great-grandfather：高曾祖父 Great-great-great-great-grandfather：太祖父

图 5 不同时期儿童可以追溯到特定祖先的比例

这些构建出的父系家谱可以帮助我们衡量父系血亲网络。另外，现存的最早一册户口册中记录的家庭成员内部的关系还可以帮助我们推断出源于同一在最早一册户口册编造之前就死亡的祖先的族人，从而延伸这一家谱的时间跨度。比如，那些在最早户口册中记录为堂兄弟的男性有共同的祖父，他们的儿子则有共同的曾祖父，孙子有共同的高祖父。而那些互为第二代表兄弟的男性则有相同的曾祖父等等。因此，即使在最初的户口册中，我们也能追溯四分之三的个人的第三代表兄弟的亲属关系，如果他们有这样的亲属的话。对于 20 世纪初户口册的记录中，三分之二的个人都有可能确定其第六代的堂兄弟关系。

通过这些联系，我们分析时便可以按照同一祖先的后代进行分组，来研究两代人间的特点或者某一特定结果的传承性和代际相关度。2007 年初我们对 CMGPD_LN 的部分数据进行计算，发现通过共同祖先分类，追溯到最初有效的户口册中出现的祖先，可以将 139,435 位记录的男性划分成 11,832 个世系。世系间的大小有很大的区别，其中规模最大的前百分之一中的男性占所有男性的百分之十五，最大的前百分之十的家族中的男性占所有男性的百分之五十七，最大的头百分之二十五的家族中的男性占所有男性的百分之八十。我们还对那些在最初户口册编造前便已去世的祖先的后代进行分组，得到了 976 个世系，每一个世系平均有 142 名男性。 其中最大的包含有 2545 名男性。前百分之一最大的世系囊括了 11%的男性，前百分之十则囊括约一半的男性。

2.C 局限性

2.C.i 对儿童的遗漏

由于户口册只记录那些在每三年户口册更新时在世的个人和上一次登记中出现、此时过世的个人，因此户口册中遗漏了许多夭折的婴儿和儿童。从这个角度看，户口册更像是一个每三年的人口普查，如同一个三年期的跨区域人口情况快照。户口册并不专门把出生作为一个事件来记录，这与欧洲国家的人口登记所不同。如果一个新生的孩子在下一次户口册更新日期前死亡，那么他们不会有任何记录，在整个数据库中不会出现任何相关的信息。许多父母一般都等到孩子长到 5 岁以后才登记，再者，由于很多户口册遗失，一些可能在幼年时期死亡的个人也被遗漏了。

户口册同样把大部分的女孩遗漏了。在 1840 年前，只有在沈阳附近的地区的户口册才一定规模地记录女孩。即使在那些户口册中，孩子的性别比例仍然是很不平衡。1840 年之后，就再也没有一定规模的女孩记录出现在户口册中了。

因此，通过对孩子的记录推断出生事件并估算生育率而得出的结果是不完整的。还需要进一步根据存活对婴儿的性别比例及男婴或者男童的死亡率的推断来对生育率进行修正。Lee and Campbell (1997, 65-70)提出了一个包括对婴儿或幼年期死亡率修正的估计。在解读事件史分析得到的生育率结果时，还需要考虑到生育率中存在的看似明显的差别很有可能是因为婴幼儿死亡率的不同，或者只是因为记录女童的意愿的不同。

2.C.ii 遗失的户口册

部分原始户口册因损毁、错放、残缺等情况难以使用，造成部分登记遗失，使得相邻户口册间的时间间隔超过了三年。如果两册户口册之间仅一册户口册遗失，那么 CMGPD_LN 中存在的两次登记就会有六年的间隔。如果有连续两册户口册遗失，那么便有九年的间隔。由于火灾事故，自 1888 到 1903 年间没有任何留存的记录，即所有行政人口 1891、1894、1897、1900 的户口册都遗失了。

如果一个遗失的户口册中刚好记录了某些个人死亡、出嫁以及其他离开方式的事件，那么在比较 CMGPD-LN 中两个现存户口册时，就会发现这些人似乎是消失了一般。例如，如果关于某一行政人口 1795 年和 1801 年的两次户口册现存，而 1798 年的遗失了，假使某一个遗孀 1795 年时仍活着，1798 年时被登记为已去世，1801 年户口册中已没有记录，这种情形下我们便无法确定她究竟是去世还是改嫁出户了。因此，我们用诸如 NEXT_DIE、NEXT_MARRY、NEXT_REMARRY 等变量对个人的离开进行事件史分析时，离开事件应该仅用那些事件发生先后两次户口册都完整的观察量。

当然，如果时间点上多个连续的户口册均遗失的话，还有一些个人的记录会被 CMGPD-LN 完全遗漏，即此人的出现和离开均发生在遗失户口册所覆盖的时间段。比如说，1801、1813 的两次户口册现存，而期间的 1804、1807、1810 均遗失。某个男孩本应该从 1804 年户口册开始出现，1807 年也有记录，而 1810 年时记录其死亡，则 CMGPD-LN 中便没有

他的任何相关记录。类似的情况，某个嫁入的妻子如果在 1804 年户口册中出现，1807 年记录其死亡，她也同样未被数据库所涵盖。

2.C.iii 事件的发生时间

对于个人退出事件，户口册只叙述了事件发生在哪一个三年的时间段中，而没有具体的日期。在数据库中，通过标识变量 DIED、MARRIED_OUT、REMARRIED_OUT 所确定的死亡或外嫁等离开事件，原户口册的注解表明事件是发生在 此次登记之前的三年间。NEXT_DIE、以及对女性而言 NEXT_MARRY 和 NEXT_REMARRY 等变量则基于对个人下一次有效户口册中“离开”而添加的注解。如上所言，如果之前的一次户口册缺失，而某些个人的死亡或其他原因的离开刚好发生，那么 CMGPD-LN 中，这些个人则似乎在两次登记之间的时间段消失了。只有这些事件的发生刚好记录在当前户口册中时，标识变量才会被设置为 1。

CMGPD-LN 原户口册并没有像标注死亡或外嫁事件一样特别标注个人嫁入某一户的事件。因此与其相关的一些变量主要是编码员和专门软件通过比较个人在每一次户口册中的状态而构建的。比如某位男性在头一次户口册中没有记录其配偶，而在之后户口册中有他的配偶信息，那么程序即在他的头一次户口册中把 NEXT_MARRY 变量设为 1。当然如果在他的先前婚姻情况中发现其丧偶，程序则会把 NEXT_REMARRY 变量改为 1。NEXT_MARRY 变量和 NEXT_REMARRY 变量的构建对于男性和女性是有不同的：对于女性个人而言，NEXT_MARRY 和 NEXT_REMARRY 变量只有在临近的下一次登记注释中明确说明外嫁时才会被设为 1；而对于男性而言，只要在当前户口册中显示其未婚或丧偶，而在之后的某个户口册中有其配偶的信息，不管这期间有多少次户口册缺失，这些变量都会被设置为 1。

当研究老龄人口死亡率时应当十分注意，因为一些个人没有死亡记录，而且他们无限地被记录下去。这些个人数量很少，影响不到成年人的死亡率，但是从 75 岁开始，因为其他人都已经去世了，他们在仍然存活人口中占相当大的比重。分母因此就比实际的分母大。这些人中的许多但不是全部在他们生命中的某一刻被记录为 ABSCONDED 。具体见对 NEXT_DIE 和 ABSCONDED 的讨论。

另外对于 NEXT_BOYS 和 NEXT_GIRLS 这类变量记录了与某个男性或女性相关的在一定时间框架内的生育事件，也是通过程序分析而构建出的，即根据某子女头一次在户口册中出现时登记的年龄计算出生育日期。大多数情况下，孩子都是在 5 到 6 岁的时候才开始被记录到户口册中，因此此类变量并不是基于某一个孩子第一次出现在户口册的日期。例如，对于一个男孩在 1816 年的户口册中记录其为 5 周岁，对应其的出生年份则在 1812，所以在 1810 年的户口册中，他的父母亲的 NEXT_BOYS 变量就会被设置为 1。如果这对父母在 1813 年至 1816 年间再没有其他的生育事件了，那么在 1813 年的登记中，NEXT_BOYS 变量又会恢复到 0。换句话说，NEXT_BOYS 和 NEXT_GIRLS 变量的值并不对应于其后的户口册中是否有子女存在的信息。

2.C.iv 覆盖面与代表性

从严格的统计学意义上来说，CMGPD-LN 中所登记的人口并不能代表中国人总体，也不能代表辽东地区的人口。如前文中所说，用来组成 CMGPD-LN 的户口册只记录了生活在

这一地区的部分人口。此外，还有一些包含在内务府户口册中但未被 CMGPD-LN 收录的人口（见附录 B）。更重要的是，在该地区，还有一些临近社区的居民没有被纳入人口登记系统中。而且，在一些或者是很多户口册覆盖的社区中也生活着未被盛京内务府登记系统记录的居民。其中一些人可能纳入在其他登记系统中，另一些人则没有被纳入任何登记系统中；因为他们耕种私人土地，而由于某种原因又没有和任何国家机构有联系比如保甲系统。

然而，从涵盖的区域来说，数据库中的人口也是有一定代表性的，因为 CMGPD-LN 集合了许多处于不同环境下的社区，包括有沿海的渔民社区、耕作在辽宁冲积平原地区的农业社区、位于辽东北部山区的社区以及其他许多有着更专门生产方式的社区，这些部分组合在一起就可以作为具有这一地区或者说整个辽东地区典型性的样本。因此，在下文的很多分析中，我们把辽东的北部、中部、中南部以及南部区别开来，如地图 2.1 中所示。故此，尽管得出的一些对相关系数及其他参数的估计并不适用于整个中国，辽宁甚至辽东地区。从 CMGPD-LN 中不同环境下得到的趋势对于研究中国其他地区至少有一定启发性。

2.D 应用实例

我们利用 CMGPD-LN 早先的版本做的分析已经对了解中国的社会组织、人口学以及历史人口统计学这一领域本身做出了重要贡献。例如，从 CMGPD-LN 中分析得到的结果帮助我们重新理解历史上的中国生育行为中的主动控制生育行为(Lee, Campbell, and Tan 1992)，从而推翻长期存在的马尔萨斯的论断——即控制旧中国人口基数的主要因素是死亡率(Lee and Wang 1999)。

对 CMGPD-LN 的分析还揭示了社会经济的分层、家庭组织、中国历史人口统计等之间的关系(Lee and Campbell 1997)。最近，CMGPD-LN 还是一个国际化的比较研究项目——古代家庭组织在应对经济危机时对调节死亡率和出生率的作用 (Bengtsson, Campbell, and Lee 2004; Tsuya, Wang, Alter, and Lee 2010)——所利用的几个数据库中的一个—。这一研究结果挑战了长期以来的学界对于欧洲与亚洲之间，家庭在应对经济危机时发挥的不同作用的看法。

在此，我们强调了一些重要的发现，它们均体现了 CMGPD-LN 有被其他研究者进一步深入研究相关课题的潜力。我们集中展现通过分析数据库中得出的社会和人口统计学结果是足以信赖的，个人、家庭、社区的特征是丰富并且多样的，因此对它的进一步分析将最终帮助我们认识当下类似的问题。同时，我们还将指出对于某一结果的分析研究是如何帮助认识当前存在的争议性课题。在最开始的部分，我们将展示如何用这些数据来研究人群健康状况与死亡率的决定因子——这也是与当下最直接相关的课题，接着我们会继续展示 CMGPD-LN 对加强我们对许多社会、行为科学领域现象的理解的潜在应用价值。

2.D.i 死亡率

应用 CMGPD-LN 进行的研究证实了其广泛使用于研究人口的死亡趋势、模型及其决定因素。早期的分析表明 CMGPD-LN 中与死亡模型相关的年龄和性别与历史上亚洲其他地区人群所观察的情形一致，即除了婴儿期、童年期间或者是极其高龄人群，对于一般人群的死亡事件的记录是可信的(Lee and Campbell 1997, 58-82 Lee et al.1992, 1995)。其后的分析又表

明，死亡风险系统地随着个体特征、家庭及社区环境变化。而且这一差别模型揭示了不同的家庭和社会组织结构(Campbell and Lee 1996, 2000, 2002b, 2004).

然而，这套数据用于分析死亡率的潜力仍然有待发掘。由于这套数据提供了个人完整的生命史，这些数据便是用来研究个人早年生活状况对于其晚年死亡情况影响的最佳选择，如Bengtsson and Lindstrom (2000, 2003) 和 Costa (2000)等进行的研究课题。我们也已经确认了这些数据进行此类分析的适用性，并证实童年时期失去一方父母对于成年时期的死亡率有较大的影响(Campbell and Lee 2006)。但是，鉴于这套数据拥有如此丰富的可以作为解释变量的个人、家庭及社区的特征的信息，许多更深入的研究都有待我们去做。这一数据库还可以用于研究老龄化问题的关键议题，例如，诸如家庭中存在成年子女等一些理论上非常重要的因素对老年人死亡率的影响。早先的研究中，我们还证实了数据库具有足够的样本量来揭示亲属构成和其他一些家庭相关的特征对于死亡率的影响，以及在经济冲击情形下死亡率波动的敏感性(Campbell and Lee 2000a,2004; Lee and Campbell 2005)。

2.D.ii 健康状况

CMGPD-LN 是为数不多的同时具有历史性、长时段、并细化至个人层面的数据库，它在提供死亡结果的同时，还提供了个人的健康状况。早期的户口册提供了男性的详细的残疾状况，从较为常见的痨病、中风、精神疾病、智障、失明到比较少见的创伤如老虎咬伤等等。其后的户口册虽不再那样详细，但仍然有注明了因慢性疾病而被豁免赋役的男子。我们已经证实被标注为残疾的状况预示了高死亡率，而且，当具体残疾类型的信息被提供的情况下，残疾的程度越严重，死亡率也越高，例如，那些被注明有神经疾病、痨病、中风等严重疾病的人比起那些只是跛脚或兔唇的人死亡率高出一截(Campbell and Lee 2003a)。

同目前已经成为重要的研究社会经济成就、慢性疾病、死亡率之间的关系的重要历史数据的美国内战时期的联军数据库（Union Army Study 或者 UAS）（Costa 2000）相比，辽宁多代人口数据库的优势在于其连续、详尽的记录了家庭、社区的环境，以及完整地包涵了个人的生命史。这些优势便于我们研究多个和当下密切相关而 UAS 所无法解决的问题。这些问题包括父母的特质及童年时期家庭情况是如何影响被个人在成年时被户口册记录为残疾的概率；残疾记录对个人在其后的婚姻、迁移、生育、死亡等事件上的影响又是什么；家庭特征对上述影响的调节如何，以及残疾记录对残疾者的可能照顾者——配偶、年长的父母、成年子女等——的人口学和社会经济后果将会有怎样的影响。

2.D.iii 生育

CMGPD-LN 同样也是研究人口繁衍的绝好材料，但是对于研究生育率不够理想。尽管户口册忽略了许多在婴儿期或儿童期夭折的男孩和绝大多数未婚的女孩，但所有家庭中存活下来的男孩的记录都是完备的，1840 年以前在沈阳附近村落中存活的女孩记录也比较完整(Lee and Campbell 1992, 1997;Tsuya, Wang, Alter, and Lee 2010)。因此，虽然不可能以每一次出生事件来衡量和研究生育率，我们可以考虑用存活的男孩数量及个别时期个别区域内存活的女孩数量来衡量生育率。从我们关于经济状况、个人特质、家庭情况等因素对生育的影响的分析结果来看，这一方法确实有助于我们了解中国人口的动态变化以及在比较视角中研究人口的动态变化(Campbell and Lee 2001b; Lee and Campbell 1992,1997, 2005; Tsuya, Wang, Alter, and Lee 2010)。

数据库中关于血亲网络的信息还使 CMGPD-LN 适用于分析遗传学中的一个关键方向：人口组成与其生殖适度差异的关系等。CMGPD-LN 数据不仅在血亲层面而且在社区层面涵盖了这些细节，使我们有可能研究同代血亲或是多代血亲间在生育事宜上的相关联情况，并评价这一关联对社区人群构成的影响。我们之前的研究也已经展示了，个人的不仅是近亲还包括远亲的特征是怎样影响到生育结果(Campbell and Lee 2001a, 2008a and b)。即使如此，还有太多的问题没有被深入探讨。这些数据还可以用来衡量某一时段内同族亲属社区与异族亲属社区各自的趋势，和这一趋势与在不同血亲团体中成功生育事件数量不同的联系。具体来说，我们可以去探讨同姓村——即所有的村民有同样的姓——的出现和异姓村——有多种姓氏存在——能够维持下来的内在规律。

2.D.iv 迁移

CMGPD-LN 中的一大特色之处在于其可以作为研究前工业化时代环境下的人口迁移行为，这主要得益于户口册中不间断地登记人口的迁入与实际居住民的情况。绝大多数其他的记载历史人口的数据库都是以社区或是以现今居住地的不同进行划分，这就意味着在研究人口迁移行为时需要大量的人力来将在不同社群中所包涵的同一个个人进行定位和建立记录间的关联。例如在 LMGP 中所涉及的 29 个群体，和我们数据库所不同的地方，是以世系制度的方式来定义和组织成员的加入事件。这种设计造成的结果，即哪怕一些个人已经迁出到另一社群中，他们仍然会被保留原来的户籍，且会继续以同样的方式记录在原来的户口册中。因此，每一个个人从他最初相关联的地域到逐渐散布开的整个过程都是可被追溯 的。CMGPD-LN 中的所有记录加在一起包含了约有超过 21000 人次被记载已经迁移到其他地区，这些地区包括从辽东北部、中部、中南部到南部。当然还有更多的人次被追溯到在区域内部进行了迁移。我们已经证实 CMGPD-LN 中的数据还可能用于研究社群、家庭及个人的不同特质是如何影响人们决定迁移到其他群体或是彻底离开的几率 (Campbell and Lee 2001b)。

使用这一数据库来研究人群的迁移行为，时间跨度从先民最初在辽宁地区定居下来（或者稍后一段时间）一直到二十世纪初，还有很大的空间等待我们去挖掘。这些数据可以帮助我们预测估计“目的地——选择”模型，使用包含潜在目的地的特质及其与作为解释变量的个人和家庭特质之间的交互项的条件逻辑斯蒂回归 (Mare and Bruch 2003)。另外，因为这些数据可以被用来研究通过迁移散布到多个社区的血亲团体，即我们可以做出直接的判断：这一行为是否可以作为一种抵御风险的策略？(Stark and Lehvari 1982)。数据的容量越大，细节越多，对于那些研究方法论的科学家而言——即使他们本身对前工业化时代的人口群体并无兴趣——这些资料却可以提供一个测试平台来检查他们新提出的一些在研究人口迁移领域的定量方法。

2.D.v 社会经济地位

在众多的用来研究社会阶层、社会流动性的人口数据库中，CMGPD-LN 同样因其细致的有关社会经济成就的记载而与众不同，这不仅有个人的成就，同样也包括了他们整个血亲网络。在大多数研究中，由于受制于有限的数据，都只考虑了个人父母的特质情况来作为社会经济成就的决定因子，却忽略了其他一些关系很亲近——祖父母、叔父、表兄弟等血亲情况。相反的，在 CMGPD-LN 中，得到的结果不仅考虑了父子关系，还考虑了兄弟姊妹、表

兄弟和其他更为广泛的家庭内部群体，甚至会考虑到这些人之间不同的地理距离对结果所产生的影响。在我们的研究中早已经揭示了这一联系的存在，并且十分重要(Campbell and Lee 2001b, 2003b, 2008a)。人事上的结果不仅会因近亲属的社会经济状态而受影响，那些远亲，包括生活在其他家庭中的亲属也都会影响这一结果。

我们最近的研究也集中在这一方面，即家庭内部成员成就之间的联系，包括维度、来源、影响等各方面，特别是从长期角度看(Campbell and Lee 2007)，当然，这些数据的大部分价值还远未被挖掘。比方说，数据中详细的有关个人社会经济方面成就的记录以及在代际内与代际间发生的人事行为对于研究它们是如何塑造人群组成具有非常的意义。而如今，这类研究社会经济流动性、人群差异和人口组成之间相互作用的课题都受限于目前各类数据的局限性，需要预测和模拟(Mare and Maralani 2006; Musick and Mare 2004)，因此，CMGPD-LN 的广泛容纳性便可以提供对这些重要过程进行直接研究的机会。

2.D.vi 家庭，亲属关系与社区

CMGPD-LN 另一个最鲜明的特色即它同时也细致记录了血亲网络和社区信息。这种同时将分布在不同社区的血亲族群以及由不同血亲群体所组成的社区展现在我们眼前，帮助我们解开并衡量亲属及社区对个人人事、社会层面产生的影响结果，同时包括对二者相互作用的影响。我们最早的研究也已证实数据库对于建立多层次模型来识别亲属和社区对个人的独立影响及共同作用影响这类话题的良好适用性(Campbell and Lee 2008b)。由于这类需要利用户、血亲团体、村庄的集群的方法发展迅速——得益于统计理论和计算资源上的不断进展——它们极可能是非常有前景的应用领域。即使在前工业化时代亲属、社区在塑造个人形成过程中的影响这类问题并非研究的重点，但这些数据仍旧可以作为相关估计评价方法的一个测试平台。

2.D.vii 空间分析

CMGPD-LN 的另一个未被开发的领域则是它精确的记录了绝大多数人口所属社区的地理位置。目前，评价不同地域在人口和社会经济成就方面的区别还仅限于地区间的粗略整体比较(Campbell and Lee 2004; Lee and Campbell 2005)。这些数据提供了许多应用 GIS 技术来定位，或是进一步理想地分析出地区差异对人口和社会经济方面结果产生影响的可能性。最近在 GIS 技术领域的进展可以进一步的将当下数据与历史数据同时用来分析人口和社会经济结果。现今的数据会包括海拔、土壤、植物、森林覆盖率、地形、降水量和温度等细节，这些数据完全可以替代历史上某一时期该地区的相关信息。而历史数据则会包括当地人口密度、临近的陆路和河流以及主要的城镇和城市。这一类研究的成果可以作为相近的 如 Gutmann、Sylvester 等人对北美大平原地区——堪萨斯州——人口进行研究的有力补充(Sylvester 2009)。

2.D.viii 制度背景

CMGPD-LN 中同样未被深入探索的课题还有制度和种族构成等时代背景又是如何影响到个人诸如婚姻、生育、疾病、死亡、迁移行为以及如何影响到其所属家庭和宗族这类背景。在第四部分中我们已经说明，CMGPD-LN 中所含人口至少包括了四种民族（汉族、朝鲜族、满族和蒙古族），行政归属上分三类（正身旗人、特殊义务旗人、庄园旗人），身份上分为两

类（正身、奴仆）。身份种类和背景的划分是严格的。正身旗人可以寻求国家提供的岗位，这些岗位的来源包括地方和八旗以及科举等系统，而旗人奴仆却不能。这些不同界限的划分显然会对个人的生活产生重要影响，很多研究制度与社会领域的专家都认同这一观点，但遗憾的是，目前为止只有极少数经验性的研究探索了这一因素对人口和社会层面的真实影响，而运用 CMGPD-LN 便是一个理想的起点。

3 CMGPD-LN 中基本变量说明

3.A 组织

CMGPD-LN 分几个文件发布数据，使用者可根据需要分别下载，再将相关的文件合并起来。基本版数据、分析版数据、血亲版数据和限制版数据由从原始户口册的内容创建的三年一次的个人观察值组成。利用相同 RECORD_NUMBER 的值，可以将同一个人同一年的观察值匹配起来，这也是将不同版本数据合并、建立新分析文件的基础。

CMGPD-LN 附加版数据提供了更多有关户口册中地位记录或涵盖区域制度背景数据的信息。它们有些是基于文件中相关变量特定的值，被合并成了个人层面的数据。

3.B 处理缺失值

我们对于所有变量的缺失值都进行了统一定义。缺失值变量共分为两类。

我们使用-99 来指代这种情况：原始数据中某类观察值通常应记录这一变量所包含的信息，但是在这一特定的观察值中相应信息缺失。一个普遍的例子是，对于 AGE_IN_SUI 这个变量，在原始数据据大多数观察值都会记录一个年龄，但一部分观察值却没有记录年龄。若观察值被标明自上次登记后已死亡或是其他原因的退出，户口册中很有可能忽略掉它们的年龄信息。这些原始数据没有年龄信息的观察值 AGE_IN_SUI 的值即为-99。

我们使用-98 来指代另一种情况：原始数据中某类观测值通常不会记录这一变量所包含的信息。一个例子是女性观测值中的各种行政地位变量。因为原始数据户口册一般不会记录女性的行政地位，所以这些变量在所有女性中均为-98。同样，在 1789 年之前，户口册中并没有区分居住户，所以很多涉及到居住户的变量在 1789 年之前的所有观察值中都被设为-98。

3.C 基本版数据（DS1）

CMGPD-LN 中主要涉及三类变量：原始变量、标识变量和分析变量。原始变量是由编码员直接将原始户口册记录的信息录入或进行诠释得到的；标识(ID)变量是由计算机程序生成的，包括对记录进行长时间和多世代的连接，以促进对数据库的管理；分析变量是一些支持人口和社会分析的计数变量、事件变量和发生变量的有限集合。这些变量都由计算机程序根据户口册的内容生成。在分析版数据中有更丰富的分析变量集合。我们解释了每一个变量是如何被构建的，阐述了它的具体特征，以及使用者应该注意的该变量的特殊性和缺陷。

3.C.i 原始变量

数据编码员根据户口册户口册的内容对原始变量进行录入或赋值。其中，有些变量是标识变量，用来标明在原始记录中某一项标注是否出现，或原始记录是否标明某人的某一特定身份。我们分三部分对原始变量进行讨论：3.A.I 主要介绍一些基本的变量如年龄、家庭关系及其他一些特征。3.A.II 描述了居住区域相关的变量。3.A.III 介绍了人口学变量，如死亡、婚姻等。3.A.IV 最后介绍描述不同行政地位的变量。

每一个完整的户口册户口册都包括封面和记录页。例如图 6 和 7 显示的就是摘自辽东八旗户口册户口册中 1822 年郎家堡魏家屯汤岗子海州岳州人丁户口册的封皮与记录页。如图 6 中所示，封面上记有户口册户口册的名称，这也是我们 DATASET 变量中所记录的内容。另外还记有户口册户口册制作的年代，我们的 YEAR 变量便以此为根据。

CMGPD-LN 基本发布数据中其他变量也主要是通过户口册户口册中这样一些信息来构建：人际关系、居住地、年龄、重要的人口事件、行政地位头衔以及族长的记录等。图 7 中显示的是记录页 1789 后的基本格式，即每人一列，并有特别的标识表明新的族或户记录开始位置 1789 年之前，个人是一个连一个记录的，每一列往往会记录多个个人，另外只有族之间才有标识进行区分，户之间并没有。

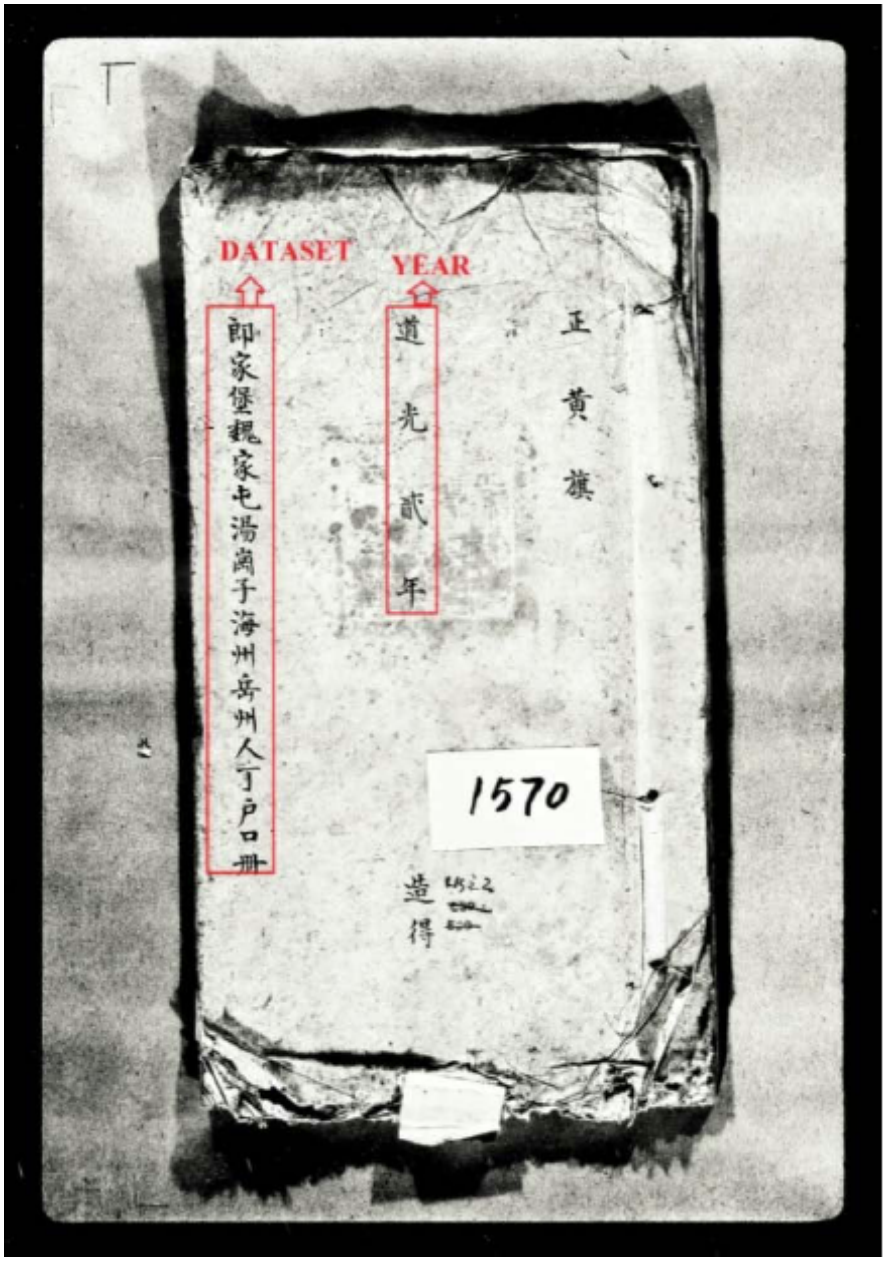


图 6 编纂于 1822 年的郎家堡魏家屯汤岗子海州岳州人丁户口册的封面

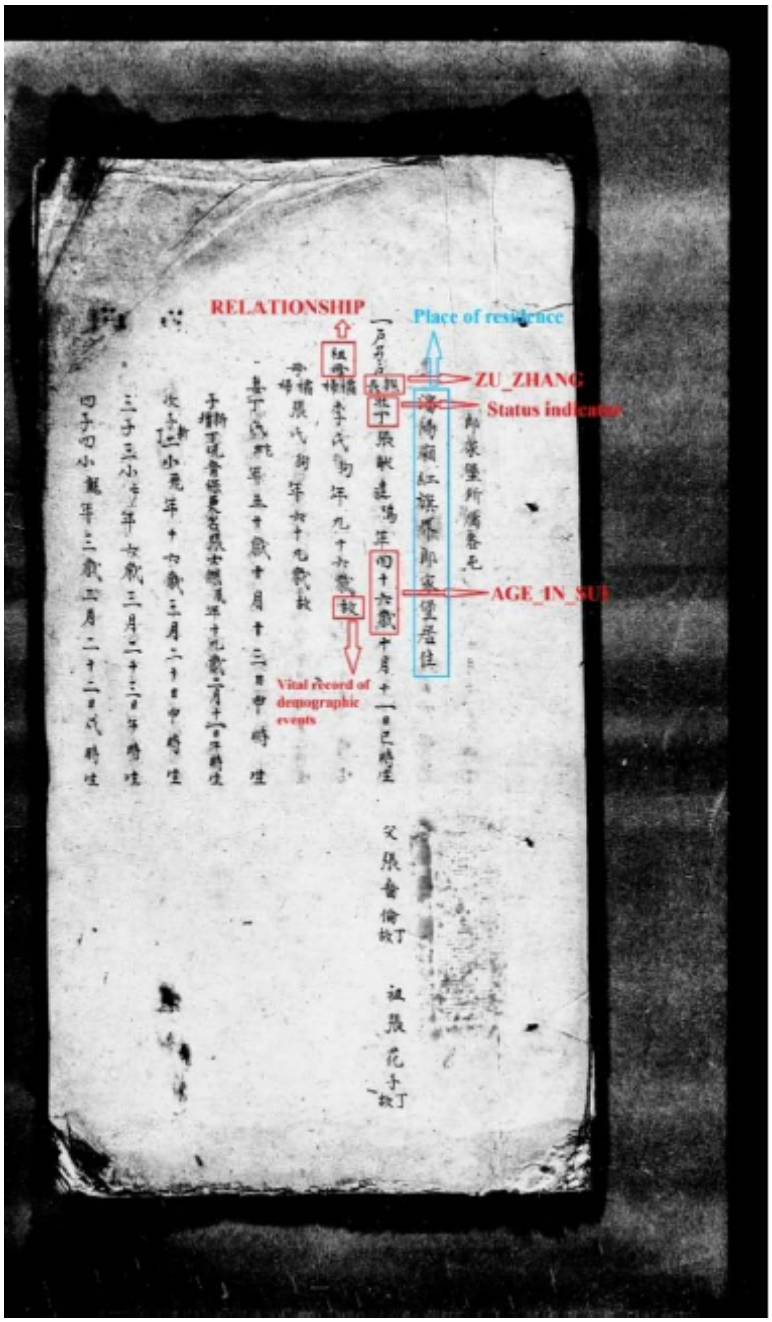


图 7 辽东八旗户口册登记页式样

3.C.i.a 基本变量

DATASET 人群数据

DATASET 标明了在各册户口册封面记录的特定旗人人群信息户口册。其数值由编码员根据户口册户口册封面上的旗人人群名称直接编码。在 1864 年以前，大部分户口册户口册以地区名称加上人群的功能作为名称。图 6 户口册是 1822 年制作户口册封面，在它左侧的第一行即写着：“郎家堡魏家屯汤岗子海州岳州人丁户口册”，这表明户口册户口册中的人口就来源于这一地区。这一人群数据的代码是 17。

1864年之前，大部分户口册用地名加上人口的功能作为户口册的名称。图6提供了1822年出产的一本户口册的封面页为例。文本第一行的左边写着：“郎家堡魏家屯汤岗子海州岳州人丁户口册 ，” 代表了这是郎家堡魏家屯汤岗子海州岳州的户口册。这支旗人数据库的编号是17。

1864 年，户口册的名称记录的更为详细了。除了地区名称人群功能外，许多户口册还把人群的类别也加入其中。如在 1864 年，人群的类别——“汉军”便增加到郎家堡魏家屯汤岗子海州岳州户口册变成了“郎家堡魏家屯汤岗子海州岳州汉军人丁户口册”。因为同时包含地点、人群分类和功能的名称可以对旗人人群进行一个全面的概括，所以各人群代码标注统一基于 1864 之后的封面名称产生。

尽管大部分户口册名称中均包括了地区名称、人群分类与功能等信息，也存在部分户口册只简单的根据所记人群的性质或功能来命名。比如表 1 中显示的 DATASET 3 的名称为“打蜜汉军人丁户口册”，这里的“打蜜”指收集蜂蜜或蜂蜜收集者，并没有记录具体的地区。另外，如 DATASET 27 名为“开档”，意为这些人是被八旗登记系统驱逐在外的人群。

因为八旗以人群而不是地区作为管理的基础，所有的数据都有一般共有的性质。首先，人群名称是由地区名称和人群功能信息组合在一起的，DATASET 1 中“道义屯汉军人丁”是地区名称，而 DATASET 3 中的“打蜜”则显示了人群功能是采集蜂蜜；其次，即使某一人群的名称是以地名命名的，但实际上这并不意味着该地区就是所有人口目前的居住地，而只代表行政登记的最初地点。在 CMGPD-LN 中所覆盖的超过 160 年间，一些从属于 CMGPD-LN 中某一人群的个人在辽东区域内迁移过多次，有些甚至迁移出辽东地区。我们的数据库中记录了许多这样的地址变动。如表 22 显示各人群共有有超过 21000 人，被记录有辽东境内不同区域的居住地址。

在本手册的第五部分对 CMGPD-LN 中的人群进行了更详细的介绍。编码本中包含了一张按人群分类的观测值分布表。这些旗人人群在社会经济背景和地位上有很大的区别。那些对人群间比较，人群类型研究有兴趣的数据使用者者可以阅读第五部分中对每一个人群数据进行的讨论，以对个人群的特性有更好的理解。

YEAR 年份

YEAR 记录的是该户口册制作时的公历年份。这是由数据编码员通过记录在户口册封面上的年号来推算得出的。例如在图 6 中显示的封皮上，中间竖排的一列字“道光二年”——即意味着该户口册是道光皇帝在位第二年时制作的。因为道光是在公元 1821 年登基，那么第二年则是 1822 年。

尽管 CMGPD-LN 提供了从 1749 年到 1909 年间人口学和社会经济学数据，但其数量与质量在年份间是有差异的。一般来说，辽东八旗的人口登记在 1789 年前的记录相对简略，只记录到族的层面；而 1789 年之后，登记变得更为正式，登记的最小单位也变成了户。除此之外，早期的大部分数据均已缺失。总体来说，在 1749-1789 四十年间留存的记录数只占全部记录的 8.09%。表 1 说明 29 个人口数据库中有 5 个是从 1789 年开始；还有另外 5 个数据库只是从十九世纪中后期才开始，这一问题在第五部分中也有讨论。

RELATIONSHIP 关系

自 1789 年开始，如图 2 和 7 中所显示的，RELATIONSHIP 变量指代所记录的个人观察值与户主间的关系。编写者用一串数字和字母来描述这一关系：每一个数字代表个人在户口册中的出生顺序，而每一个字母代表与户主间的关系。1789 年之前，如图 1 所示，RELATIONSHIP 代表的是个人与户族族长间的关系。所以 1789 年前的 RELATIONSHIP 编码相对比较复杂，因为一般族会包括许多族长的远房亲戚，甚至包括第二代表兄弟甚至第三代。

在 1789 之前,正如图 1 中显示，RELATIONSHIP 实际上代表着一个人与户主或者族长的关系。给 1789 年之前的关系变量编码是十分复杂的，因为一个户经常包括大量户主的远亲，包括第二代甚至第三代表亲。

表 2 中列出了字符变量中所用字母的意义。e 是指户主或族长，并总是单独出现；1789 年以前，e 指族长，而从 1789 年开始，e 只说明户主身份。m 一般指示户主的母亲。剩下其他一些代码加在一起表明从户主，下到个人的关系。数字代表个人在户口册中的年长顺序，其后数字则是其家庭关系的内容。比如，f2yb3s 就代表户主的父亲的第二个弟弟的第三个儿子，即也是户主的第一代表兄弟。同理，f2yb3sw 就指代该男子的妻子。此外，w 单独出现则指代户主的妻子，f 单独出现指代户主的父亲。

表 2 RELATIONSHIP 编码解析

编码	关系
E	本人或户主
W	妻子
Q	妾
M	母亲
F	父亲
B	兄弟
Z	姐妹
S	儿子
D	女儿
O	年长的
Y	年轻的

RELATIONSHIP 变量是 CMGPD-LN 数据库的一大优势。它是构建诸如 FATHER_ID、MOTHER_ID、 HUSBAND_ID 和 WIFE_ID 等标识变量的基础。程序在构建父亲标识的时候会通过将某一个人关系代码的最后关系字符去掉，然后搜索满足之前部分的个人来找寻这个人的父母亲。例如，如果我们要定位 f2yb3s 这一个人的父亲，程序便搜索 f2yb，若定位其母亲则搜索 f2ybw。如果需要建立 f2ybs3w 个人与其丈夫的关系，程序便搜索 f2yb3s。以这些标识作为基础，才能有更复杂的世系与亲属关系变量，如在第二部分中提到过的祖先和亲属变量，这些变量按照世系对个人进行划分的同时度量某一特定亲属关系的特质。

RELATIONSHIP 在一户之内可以单一标识某一个人。换句话说，排除那些因字迹模糊或是录入错误的情形，在某一户家庭中，所有登记在册的个人都应有不同的关系变量值，并

且不会重复 RELATIONSHIP 变量的这种唯一性，极大的便利了程序自动进行关联。

个人通常按照原始户口册中其与户主的亲缘关系来安排顺序，因此在 CMGPD-LN 中，头领排在第一位，其后是他的母亲（很少情况下是父亲），再接下来是女方或男方的祖父母，然后是他的妻子、子女、孙子女，再其后则是叔父以及叔父的孩子、孙子女、曾孙子女，接下来还有表叔父（其祖父的兄弟）和他的后代等等。

当然，使用者需要的是，这个变量在 1789 年之前并不能用来分析家庭结构。如前所述，因为在 1789 年以前，八旗户口册是以族为单位对居民进行划分，每一个个人都是通过其与族长的关系来被索引。因此，当该变量用来分析居住人口特质相关议题时，对于在 1789 年以前的观察值应排除在外。

户主的父亲极少情况下会指一位健在的个人，因为一般情形下，男性除非死亡，会一直保持其的户主头衔。通常如果列出了户主的父亲，一般他都是前任户主而自上一次户口册后去世了，会被标记为“故”或“去世”。

由于部分数据的字迹模糊，在 1513357 条记录中有 36 条观察值的 RELATIONSHIP 变量的值为缺省。

DISABILITY_CODE 残疾代码

DISABILITY_CODE 是一个表示残疾类型的数字代码，输入时由编码员设置。利用它，可以从残疾变量补充说明中合并 DATASET 和 DISABILITY_CODE，自动检索户口册中残疾类型的拼音。

在 POSITION_CODE 和 POSITION_2_CODE 中也记录了残疾状态，这是标识变量 DISABLED 构建的基础。使用者可以根据不同 DATASET 和 DISABILITY_CODE 的组合手动地在附件的编码册中查找原始汉字记录的残疾名称。

当编码员输入一个数据人群的户口册时，他们将遇到的第一个残疾类型的 DISABILITY_CODE 设置为 1，并将原始汉字和汉语拼音分别记录下来。下一个遇到的残疾类型被设置为 2，以此类推。DISABILITY_CODE 的设置每遇到一个新的 DATASET 就从 1 重新开始，所以相同 DISABILITY_CODE 的值在不同 DATASET 中可能指的是不同的残疾类型。同样的，相同残疾类型在不同行政单位的人口中可能有不同 DISABILITY_CODE 的值。同时确定 DATASET 和 DISABILITY_CODE 才可以指向一个具体的残疾类型。

只有男性才有 DISABILITY_CODE。户口册中拥有具体残疾类型记录的男性数量随着时间迅速下降。使用者利用这个变量或相关变量如 DISABLED 时，最好查看下这个变量随着时间趋势、DATASET 和其它测量变量的分布状况，以便从中选取最恰当的数据子集用于分析。

POSITION_CODE 地位代码

POSITION_CODE 是记录男性原始社会地位的数字代码，从户口册中录入信息时由编码员设置。当编码员输入一个 DATASET 的户口册时，他们将遇到的第一个官职或地位的

POSITION_CODE 设置为 1，并将原始汉字和汉语拼音分别记录下来。他们将下一个遇到的官职或地位设置为 2，以此类推。POSITION_CODE 的设置每遇到一个新的 DATASET 就从 1 重新开始，所以相同 POSITION_CODE 的值在不同 DATASET 中可能指的是不同的官职或地位。同样的，相同官职或地位在不同行政单位的人口中可能有不同 POSITION_CODE 的值。同时确定 DATASET 和 POSITION_CODE 才可以指向一个具体的官职或地位。分析版数据中有关官职和地位的标识变量就是基于原始户口册的地位信息构建的。

通过 DATASET 和 POSITION_CODE，用户可以检索具体地位的拼音或汉字。通过合并 DATASET 和 POSITION_CODE，在职位变量附属说明中可以检索到拼音，原始汉字字符可以在附件中的编码册找到。

随着时间和 DATASET 的变化，行政官职和地位的分布存在显著的变化。我们强烈建议使用者查看这个变量随着时间趋势、DATASET 和其它测量变量的分布状况，以便从中选取最恰当的数据子集用于分析。

POSITION_2_CODE 地位代码 2

POSITION_2_CODE 是一个附加的数值变量，用于表示男性的原始官职或地位，由编码员从户口册中录入信息时设置。在大部分情况下，POSITION_2_CODE 被用于记录族长或百总地位。在一些情况下，当一位男性有多个地位时，仅仅 POSITION_CODE 不能记录完全，POSITION_2_CODE 就被作为记录超出部分的领域。通过 POSITION_2_CODE, 使用者可以查询具体地位的拼音和汉字。通过合并 DATASET 和 POSITION_2_CODE，在职位变量附属说明中可以检索到拼音。

SEX 性别

个人的 SEX 性别本身并没有直接在原始的户口册中出现，所以该项信息主要是由编码人员从其他信息推断。在原始户口册中，有诸多条目可以显现出某一个人的性别信息，主要是人际关系、姓名、职业等。一般情形下，编码员是通过某一个人与户主的关系来推断其的性别。几乎所有出现在户口册中的女性要么是姐妹、女儿，妻子、寡妇，或者妾，所以 z、d、w 或 q 是她们 RELATIONSHIP 的最后一个字母。唯一的例外是，那些作为户主的女性 RELATIONSHIP 变量值则为 e，这种情形下，编码员依然会通过她的姓名或者是关于她已死亡或逃亡的丈夫的姓名的注释来推断其为女性。

在 CMGPD-LN 中，有 366 位个人的共 557 条观察值显示个人的性别在之前的户口册中有不同的性别。可以使用下面的 Stata 代码来查看所有这些事件：

```
sort person_id year
list dataset year record_number person_id if person_id== person_id[_n-1]&sex!=sex[n-1]
```

导致这些不一致情况的原因是在原始的户口册中，个人与户主关系的信息的不一致。比如说，某一个人在前一次登记中显示其为户主的女儿，然而在其后的表中又变成了户主的儿子。当然我们没有任何办法知道哪一个记录是正确的，因而在我们把原来图片形式的信息录入成数字格式的过程中，只能允许这些不一致的个例的存在。

GENERATION 世代

GENERATION 变量记录某一个人在当前户口册中相对于户主的世代数。对于在 1789 年以前，则指代相对于族长的世代数。

编码员通过记录中个人与户主（族长）的关系来推断其 GENERATION 变量的值。与户主属同一代人的个人世代数的值设为 3；比户主高一辈的家庭成员（如户主的父亲、母亲、叔伯、姨婶、父亲的表兄弟等等）世代数为 2；比户主低一辈的家庭成员世代数则为 4（如户主的儿子、侄子、侄女等）；其余类推。

由于 GENERATION 变量只反映当前户口册中同一户中家庭成员的世代关系，某一个人的世代信息会随着时间，随着户的整体结构改变而改变。如某一户主的儿子，1s，当其继承了户主地位后，即被标为 e 时，他的 GENERATION 变量就从 4 变成了 3。

MARITAL_STATUS 婚姻状况

MARTIAL STATUS 变量的值也是编码员根据某一个人在当前和上一次登记中，配偶或子女的出现来设定的。对于判别一位女性是未婚、已婚还是丧偶，编码者可以直接通过她们的人际关系以及是否有健在的配偶来判断。因为女儿在婚后会离开自己原来的户，所以所有仍旧有记录的、RELATIONSHIP 变量最后位是 d 的女儿都为未婚；此外，当某一个女儿被注释已经自上次登记后出嫁的话，标识变量 MARRIED_OUT 会被设置为 1，MARITAL_STATUS 被设置为已婚；再者，所有在人际关系变量最末是 w 的女性都是已婚妻子。如果此时她们的丈夫仍旧出现在户口册中，她们就属于已婚；而如果丈夫不在户口册中或是虽然出现但有注释其死亡的信息时，这些女性便属于丧偶一类。如果记录中有说明某个人自上次登记后已经改嫁，那么她的 REMARRIED_OUT 变量就被设为 1，同时她的婚姻状况也被改为已婚。

然而区分未婚男性和鳏夫却相对更加困难，需要通过对当前户口册分析以及和上次户口册相比较来得出结果。在原始表中没有专门的注释来说明某个人是已婚、未婚或是丧偶。首先，那些登记有妻子的男性自然在当时属于已婚的类别；其次，登记有妻子并且注明其死亡的男性属于鳏夫；那些在当前表中未登记有妻子，但在前次登记中有妻子的也被划为鳏夫类别；还有当前没有登记妻子，但已经有子嗣的，也认为是鳏夫。因此，从上述原则看，如果一个没有子嗣、妻子又在最早有效的户口册之前就已经去世的男性个人就极有可能被错误的判断为未婚。同样一个没有子嗣、妻子刚好在两次有效的户口册之间嫁入又很快不幸去世的男性也会被认为是未婚。

同样的，一些被列为已婚的男性可能实际上是再婚。如果他们的首次婚姻没有孩子且发生在我们可得的最早的户口册之前。或者在两个户口册之间发生并结束且没有孩子，他们继而再婚，第二次婚姻有可能被误认为他们的第一次婚姻。

对于女性而言，被认为已婚或再婚代表着她丈夫的婚姻状态，而不是她自己的。当一位女性作为妻子在一户中出现时，我们不能将她和她出生的户或者是任何她曾经作为一名妻子的户连接起来。

MARITAL_STATUS 为 1 代表着当前已婚人群，2 代表着当前未婚人群，3 代表着当前丧偶的人群，4 代表着再婚人群。曾经结过婚的人群（包括当前已婚，再婚和丧偶）可以表

示为 MARITAL_STATUS != 2。

当涉及到使用 MARITAL_STATUS 变量时，使用者需要了解在登记过程中可能存在的不一致问题。在 CMGPD-LN 中，一共有 547 位个人的 579 条观察值出现婚姻状态不一致的情况。这里所谓的不一致，是指某一个人在前一年中还被认为已婚或是丧偶，却在之后被判断为未婚。绝大多数这种情况都是由于编码人员的错误造成的。其中大部分的情形是某一份之前认为已遗失的户口册记载有某个人已婚，后来这部分内容补充进了数据库，而在其后年代的户口册中由于没有更新数据，故仍然记载其为未婚。总的来说，这种错误是极少数的，应该不会影响整体的结果，不过总有细心的用户会发现这一问题。

因为男性的 MARITAL_STATUS 基本上是根据与之前户口册中的对比信息来设置的，如果对使用者来说，区分从未结过婚的人和丧偶者、已婚者和再婚者至关重要的话，可以把分析限制在孩童时期就被记录在户口册中的男性，和婚姻记录相对完整详尽的个人。

MARTIAL_STATUS 变量可以用另外一个变量替代，这个变量是通过程序对某一个人的所有观察事件进行循环分析，计算出的反映个人是否曾经已婚的变量。如果想了解所有这类不一致的具体情况，请使用下面的 Stata 代码：

```
sort person_id year
list record_number dataset year marital_status age_in_sui died if person_id==person_id[_n-1]&marital_status<marital_status[_n-1]&(marital_status!=1&marital_status[_n-1]!=2)&marital_status!=-99&(marital_status!=3&marital_status[_n-1]!=4)
```

AGE_IN_SUI 年龄（岁）

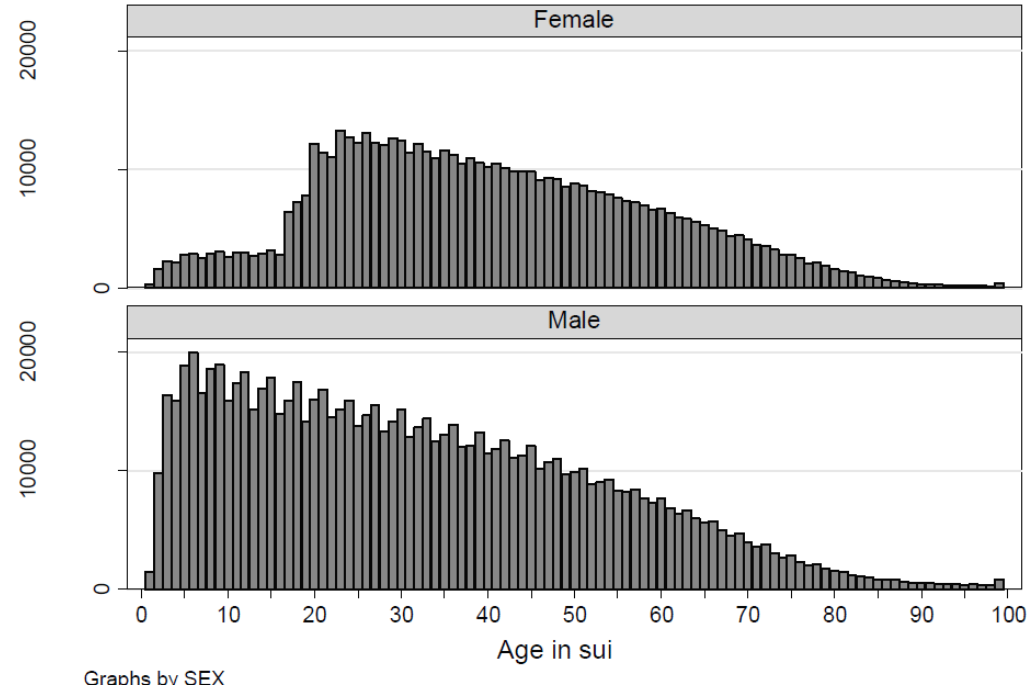
AGE_IN SUI 是对原始户口册中年龄信息的直接录入，所有辽东八旗户口册中人口年龄都是按岁来计算——中国传统的计龄方式。个人在出生时为 1 岁，每过一个阴历新年，年龄就加 1 岁。平均来看，按岁计算的年龄一般比西方计算方法大 1.5 左右。另外那些有关个人具体的生辰日期的数据较不可信，所以现在还没有合适的方法直接计算西方意义下的年龄。但为了方便同其他渠道——它们以西方模式记录年龄——得的结果相比较，一般情形下，我们在分析的时候把某一年龄段的起始年龄和结束年龄减去 1 之后来比较。例如，当我们找对应于西方年龄 5-9 岁时，就按 6-10 岁来分析。

在使用 AGE_IN_SUI 变量时，仍需注意可能存在个人年龄记录上的不一致问题。CMGPD-LN 中，我们发现有 10728 个人的 11430 个观测值的不一致年龄记录。这个问题主要是由原始户口册中的错误所导致的。解决这一办法的途径是，对每一个户口册中每一个人都通过在之前记录中 BIRTHYEAR 的信息来计算他/她的年龄。Lee and Campbell (1997, 240-242)中就评估了道义行政区人口年龄的不一致情况。

使用 AGE_IN_SUI 变量时尤其要注意：原始记录中某些个人的年龄缺失，又因为其被记载自上次登记后已迁离，而在 CMGPD-LN 中被设为-99。在某些情况下，迁离发生时的年龄会被记录，其他时候，之前原始记录上的年龄还需要通过当前户口册的年份来估计。由于这些被记录为离开的个人在原始户口册中所记载年龄的不一致性，我们强烈建议用户在做这类观察时，分析中不要纳入年龄的信息，只要简单的通过设定 AGE_IN_SUI 不等于-99，同时 PRESENT 变量值为 1 来选择观察值，就可以避免引入这类观察值。

在AGE_IN_SUI中有一些明显的三年年龄堆积,尤其是男性。图8显示了男女每年AGE_IN_SUI的比例。这种三年的堆积源于户口册每三年一册的特性，并且出生人数和同龄人群的规模每年周期性地波动。

图 8 的分布也反映了未婚女儿、男婴和男童的人数在总体上被低估了，以及已婚和丧偶女性记录的不完整。低于适宜结婚年龄的女性数量同 5 岁以下男性的数量一样明显过低。然而，20 岁及以上男女比例大致与具体年龄的死亡模式的推测一致。直到大约 40 岁，女性的死亡率才比男性的高，继而降低（见图 8）。



注：female:女性 male:男性

图 8 不同 SEX 和 AGE_IN_SUI 的样本分布

BIRTHYEAR 出生年月

这里的出生年月变量是由原始户口册中记录的年纪来计算得到的。出生年月=YEAR-AGE_IN_SUI+1.这一变量提供了一种新的计算年纪的基础，至少做为一种基于西方标准的年纪的估计。为了保证在整个库中年龄变量的一致性，即所有记录中个人的 AGE_IN_SUI 变量没有不一致情况。BIRTHYEAR 信息可以直接从个人最初记录拷贝到之后的记录中，再由 YEAR 减去它得到年龄的信息。

BAZI_YEAR 生辰八字年份

个人生肖属相的编号。一共有 12 个值。缺失值和难以辨认的值被设置为-99。编码按照动物在农历中出现的顺序：1) 鼠, 2) 牛, 3) 虎, 4) 兔, 5) 龙, 6) 蛇, 7) 马, 8) 羊, 9) 猴, 10) 鸡, 11) 狗, 12) 猪。

*在使用生辰变量(BAZI_YEAR, BAZI_MONTH, BAZI_DAY, BAZI_HOUR)时应当十分注意。*对原始户口册中生辰信息的编码十分混乱。因此 BAZI_YEAR 和其它表示此类信息的

变量在使用前都必须通过检验与 YEAR 和 DATASET 的列表来确保后续分析是在记录可靠的户口册中进行的。从 19 世纪早期开始，出生日期不正常的重复意味着记录员不再按册逐年记录信息，而是直接抄录邻近个人的信息。生辰变量 Bazi 集中代表着占卜时用的年、月、日、出生时辰。

BAZI_MONTH 生辰八字月份

记录在原始户口册中个人出生的农历月份。缺失值和难以辨认的值被设置为 -99。使用生辰变量的警告信息见BAZI_YEAR。

BAZI_DAY 生辰八字日期

记录在原始户口册中个人出生的农历日期。缺失值和难以辨认的值被设置为 -99。使用生辰变量的警告信息见BAZI_YEAR。

BAZI_HOUR 生辰八字时辰

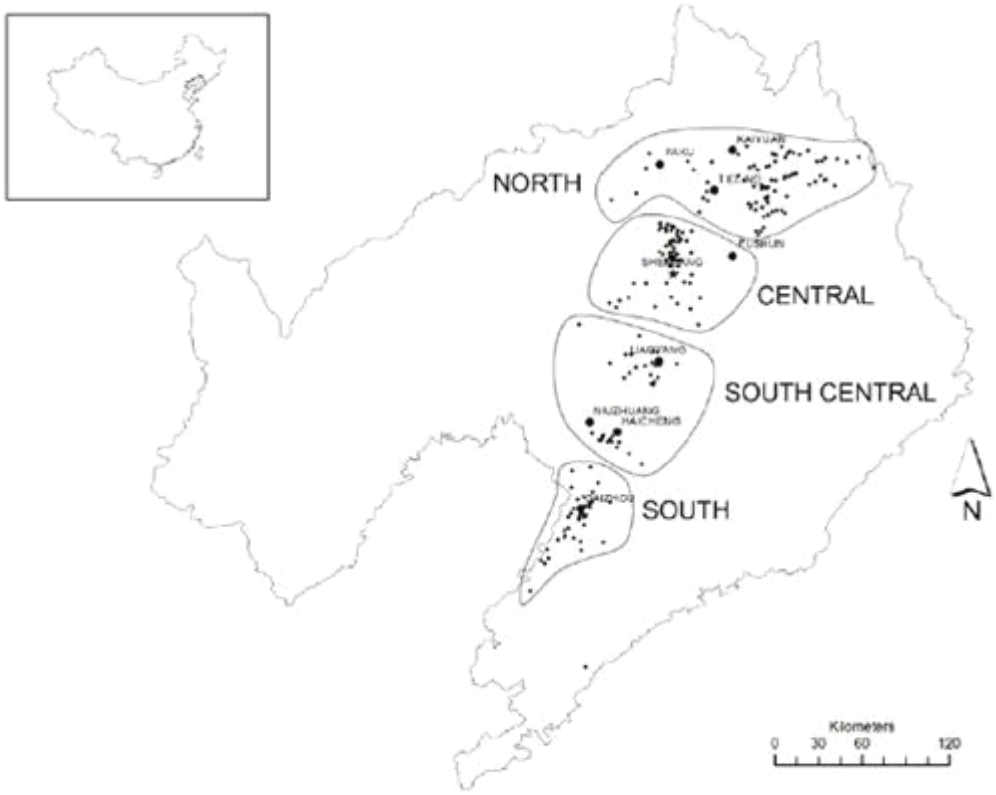
记录在原始户口册中个人出生的时辰。将一天 24 小时分为 12 个时段，按照时段编码：1) 子, 2) 丑, 3) 寅, 4) 卯, 5) 辰, 6) 巳, 7) 午, 8) 未, 9) 申, 10)酉, 11)戌,12) 亥。

3.C.i.b 居住地

八旗户口册中记录了每一户的居住地，包括从大的行政区划，到其从属小区域，再到村庄名称整个的信息。大部分记录中都明确了具体的村庄名称，只有一小部分例外。如在图 7 中显示的，居住地一行写着“沈阳镶红旗界郎家堡居住”意味着接下来的这些户居住于沈阳镶红旗界内的郎家堡村。为此，我们由这些居住地信息建立了三个变量-REGION（区域），DISTRICT（行政区划）和 UNIQUE_VILLAGE_ID（唯一村庄编号）来鉴别评价每一个个人、户的地理和社区特点。

REGION（区域）

REGION 变量记录了数据库中所有个人在辽东范围内的居住地分划。我们主要把辽东分成四个区域：辽东北部，辽东中部，辽东中南部，辽东南部。辽东北部包括开原、铁岭地区，主要是现今辽宁省东北部较为偏远的丘陵和山区农业耕地。辽东中部，包括抚顺，巨流河，沈阳，兴京等主要围绕于沈阳北部市郊，或者可以说是辽东平原等一系列地区，这些地区耕地广袤。辽东中南部则指广宁，辽阳，牛庄和熊岳等地，也属平坦的可耕地。辽东南部，即盖州，锦州，岫岩等沿海一线，尽管土地本身并不太适合耕种，这些地区却有繁荣的农业商业经济，从 1858 年天津条约中开放牛庄口岸一条就可以看出。



注： NORTH: 北部 CENTRAL: 中部 SOUTH CENTRAL: 中南部 SOUTH: 南部 Kilometers：千米

地图 2 辽东地区示意图，对应 REGION 变量中的四个区划

当原始村落有具体的行政区域和村落名字时， REGION 根据这些信息直接录入。当这些信息缺失时，我们根据可知行政区域和村落名称的样本，依据国营农场人口形态来计算 REGION。REGION 可用于不同地区之间人口的比较,但是地区间迁移的研究应该只包括原始数据中有地址记录的观测值,也就是说, UNIQUE_VILLAGE_ID 没有缺失(-99)。另外，在同一个数据库中，一个人一条记录的 REGION 与前一条记录的 REGION 不同，说明所在地区发生了改变。

DISTRICT 行政区划

DISTRICT 变量就是对户口册中居住地记录中包含的行政区划做以直接的录入。举例说明，由图 7 中所示，居住点显示村庄位于镶红旗沈阳地区界内。这里提到的历史地点的位置和当代的行政区划大致是相同的，但并不完全一致。

UNIQUE_VILLAGE_ID（唯一）村庄名称

UNIQUE VILLAGE ID 是 CMGPD-LN 中所有涉及到的人口居住村庄的名称。在 CMGPD-LN 由 DATASET 区分的各人群中村庄编码是一致的，因此那些在同一村庄而并不属同一行政人群的观察值在数据库中具有相同的 UNIQUE VILLAGE ID。这些村庄的名称则是根据原始户口册录入，在限制发行版数据中就可以看到这些信息，此外还包括这些村庄所

对应的具体地理位置的信息。当在原始户口册中村庄名字相似但不完全相同时，我们和编码员手动检查了户口册来决定它们是否指的是同一个村庄。当实际上是同一个村庄时，原始户口册中村庄的名字将会被赋予同一个 UNIQUE_VILLAGE_ID。如果不同户口册中村庄的名字只有极小的区别时，这些村庄会被认为是同一个村庄。例如，在两本连续的户口册中，用同音字代替村庄名字中的一个字，居民列表相同。有些时候，地方志或其他信息来源明确记载了一个村庄在一个时期被改了名字。

UNIQUE VILLAGE ID 变量在基本版数据中的作用是对所有观察事件按村庄进行划分，这样一来，村庄就对应结构模型（hierarchical model）中的某一层次，或者构建村级人口的概括性测量尺度。单纯使用 UNIQUE_VILLAGE_ID 变量并不能充分辨明一个有名字的村庄。整个 CMGPD-LN 中包括 699 个不同的村庄，而那些并没有记录其所属村庄，或是只有大致区域没有具体村庄的观察值，该变量的值被设为-99.

3.C.i.c 人口学事件

在八旗户口册中，某一个人的退出，比如死亡、出嫁外地、于外地收养、出逃，都会在户口册中该个人的一列记录最后明确标出，用来解释该个人在近三年推出的原因。对于某一个人出现的原因，如新生婴儿、婚嫁、收养等信息主要是基于编码员者的推算，因为在原始户口册中并没有明确记录。例如图 3.2 中所示的记录页两个个人被标注在过去三年过世。因此，根据原始记录表中对重大人口学事件的记录，辅以录入人员对前后户口册出现的变化进行分析，我们得以建立了以下这些变量来描述人群中每个个人的最重要的人口学事件。

这些在原始记录中所记载的退出通常都表示已经发生过的事件，因而相应的这些标识变量如 DIED 死亡，MARRIED_OUT 外嫁，REMARRIED_OUT 改嫁出和 ABSCONDED 出逃都不应作为分析中的因变量。此外，当某一个人被记录为退出后，年龄信息往往会被忽略，因而对于那些退出标识变量值为 1 的观察值，AGE_IN_SUI（年龄）值缺失，为-99.

至于 NEXT_DIED（下次登记中去世）这类由程序依据原始“退出变量”进一步构建的标识变量则表示该事件发生在此次和下一次登记之间的三年间，更加合适用作因变量。我们把在原始户口册就体现出的 DIED 死亡，MARRIED_OUT 出嫁，REMARRIED_OUT（改嫁离开）ABSCONDED（潜逃）等全部信息都尽可能的包括在数据库中。这些变量进一步可以用来构建新的结果变量来满足用户的特定需求。

DIED 死亡

DIED 死亡是一个虚拟变量。值 1 代表在原始户口册记载中某个个人在过去三年间去世。

使用者须注意一点，数据库中有一些个人在多册的户口册中被登记为死亡，即 DIED 变量值为 1 的情况出现了多次。因此，对于这些人的死亡结果报告是夸大的。如表 3 中显示的情况，总共包括 266091 位个人的 CMGPD_LN 中，有 944 位的死亡（故）被多次记录。实际情况是这些人的死亡信息在一次登记中被注释后，又被重复拷贝到下一次的户口册中而未被去掉，推测这多半是当初誊写人员犯得错误。显然，人不可能死了复生又再一次死亡。因此，使用者在分析数据中须注意以个人的第一次死亡记录为准。

表 3 个人被列为死亡（故）的观察值数量

个人被列为死亡的观察值数量	频数	百分比	累计百分比
0	184,184	69.22	69.22
1	80,963	30.43	99.65
2	929	0.35	99.99
3	15	0.01	100
总计	266,091	100	

MARRIED OUT 外嫁

MARRIED OUT（外嫁）同样作为一个虚拟变量来显示某一个人是否在本次记录之前的过去三年间出嫁外地（嫁到 CMGPD_LN 之外）。因为那时婚姻行为几乎全部是以男方为主体，妻子会迁到丈夫家居住，故这个变量主要是出现在家中未婚的女儿出嫁后的户籍信息中。表 4 记录的是总共有 8319 条观察值中出现 MARRIED OUT 变量值为 1 的情况，其中有 9 例的主体为男性。

表 4 不同性别的外嫁记录

	外嫁		
SEX	0	1	Total
-99	23	0	23
女性	587,820	8,310	596,130
男性	917,195	9	917,204
总计	1,505,038	8,319	1,513,357

和死亡信息的记录类似，也出现了一些女性个人有多次的外嫁记录，即也因为她们在原始户口册中被多次记载了外嫁信息。如表 5 中显示，总共 266091 女性中有 110 名被两次记录于过去三年间外嫁。虽我们将其主要分析为誊写人员的笔误，但却排除不了这样的可能：即一位失去丈夫的女性回到娘家后又再一次嫁出。现在看来，我们的数据中还未发现这类情况。

表 5 不同外嫁登记次数的人口数

个人被列为外嫁的次数	频数	百分比	累计百分比
0	257,882	96.9	96.91
		1	
1	8,099	3.04	99.96
2	110	0.04	100.00
总计	266,091	100	

REMARRIED_OUT 再次外嫁

REMARRIED_OUT 也是一个虚拟变量，显示某个寡妇是否在当次登记的过去三年间再

次外嫁。同 MARRIED_OUT 变量一样，REMARRIED_OUT 专门的显示了丧夫女性的再婚记录。一般鳏父再婚后，其配偶几乎肯定会加入到其家庭户籍中。如表 6 中显示，在 1399 例 REMARRIED_OUT 值为 1 的观察值中，只有两例为男性。对这两例情况我们还没有做研究，也很有可能只是记录错误而已。

表 6 男女再次外嫁的观测值数量

	REMARRIED_OUT		
SEX	0	1	总计
-99	23	0	23
Female	594,733	1,397	596,130
Male	917,202	2	917,204
Total	1,511,958	1,399	1,513,357

此外，如表 7 中显示，还存在 12 位个人在原始户口册中有超过两次的记录显示其再次嫁出（REMARRIED_OUT 变量仍为 1）。考虑这种情形，使用者需要在分析时排除这类重复的情况。

表 7 不同再次外嫁次数的丧偶人数

被记录再次外嫁的次数	频数	百分比	累计百分比
0	264,705	99.48	99.48
1	1,374	0.52	100.00
2	11	0.00	100.00
3	1	0.00	100.00
总计	266,091	100	

ABSCONDED 逃

该虚拟变量显示某个个人在当前记录表中是否被记录为非法潜逃。因为所有 CMGPD_LN 中所包涵的人口都隶属于皇家内务府，个人在未获得国家的允许下是不能私自离开其所属地，那些未经官方允许而离开的人都被记录为潜逃。当然这种规则只对于男性个人有效，因为只有男性个人才负有为国家劳动和服务的义务（丁）。个人的潜逃记录会一直留存直到其回归，案例才被解决，或者最终被户口册删除。根据实际情况看，一些男性在许多年甚至几十年间状态一直显示为潜逃。若某一个人潜逃后在他地死亡并未得到报告的话，他们的潜逃记录可能会永远留存下去。所以就出现了个别人潜逃后似乎活到近乎荒谬的长久年岁的情形。

在某些例子中，个人仍旧一直被当做家庭成员，并最终活到了非常大的年纪。然而，因为活到这般高龄的个人本身不多，这些被认为已潜逃、事实上已死亡但却在户口册中一直留存的个人在整个高龄人群中占据了较重比例。考虑此种情形，我们强烈建议在对高龄人群做研究分析时，须排除掉那些在当前记录中显示为已潜逃，或是曾经被记录为已潜逃而后续再没有相关记录的个人。对于女性个人，该变量都被设为无效（-98），因为她们不涉及潜逃这类情形（表 8）。

表 8 男女逃亡的观测值数量

	逃			
SEX	-98	0	1	总计
-99		23	0	0	23
女性		596,130	0	0	596,130
男性		0	890,906	26,298	917,204
总计		596,153	890,906	26,298	1,513,357

PRESENT 存活

PRESENT 存活是由计算机自动生成的虚拟变量，主要依据户口册中记录的某个人的重要人口学事件。一般情况，个人只要没有被记录为退出并认为其存在于户口册人群中，而像 DIED 死亡、MARRIED_OUT 外嫁、REMARRIED_OUT 再次外嫁和 ABSCONDED 潜逃等情况下 PRESENT 变量值为 0，数据库中也再没有其他形式的退出信息。创建此标识变量主要是便于用户数据分析时使用，即作为一种筛选观察事件的办法，来快速找到那些当前生存并未经历某些事件而退出的个人信息，或者在计算时帮助判断平均水平。在我们进行过的分析中，对象都只限于标识为存活的个人中。

3.C.i.d 行政地位

八旗户口册原本就作为行政管理的方式存在，因此在原始表中，每位男性个人名字之前都记录了其行政地位（见图 7）这里的行政地位一般包括个人雇佣从属关系和可能拥有的官方职务，以及该职位是否有效，还包括该个人在国家科举考试中所取得的头衔。总的来说，所记录的地位信息主要分为四类：官员、军人、工匠和丁（即需要对国家承担指定的劳动和服务义务的成年男性）。下面的变量都是依据这些原始信息录入得到，以便于掌握和分析这类信息。至于个人如果有一些重要的非行政性位置，比如当地商人等，这些一般不会记录到户口册中。

RETIRED 退

RETIRED 表示某一男性个人曾经有过职位但在户口册中记载为当时已退的虚拟变量。在原始记录中，退休字样后一般跟着职位信息，比如退休的佐领记为“退佐领”，退休的男丁（退丁）。所以，程序主要通过对个人原始记录中的行政地位信息进行文字分析后来构建其的 RETIRED 变量。

退休的个人一般都存在一定的年龄范围。像表 9 显示的，84%的被记录为已退休的个人的年龄都超过 51 岁。同时，不同个人退休的原职位也分布很广。约有 0.7%的个人年纪在 30 岁以下被记录为已退休，Lee and Campbell 1997 就分析了在数据库 1 中的退休人群的年龄分布情况（165-6）。

表 9 退休人员年龄分布

年龄	频数	百分比	累计百分比	
缺失值		3,574	9.9	9.9

1-10	10	0.03	9.9
11-20	23	0.1	10.0
21-30	201	0.6	10.5
31-40	603	1.7	12.2
41-50	1293	3.6	15.8
51-60	3231	9.0	24.8
61-70	16,096	44.6	69.4
71岁及以上	11,052	30.6	100.0
总计	36,083	100.0	

OLD 老

OLD 老是表明某位个人拥有某一特定地位，但在当前户口册中被记载为“年老”的伪变量。像 RETIRED 退休一样，此变量的信息也是通过对原始记录中行政地位进行文字分析后得到的。一般情况下，关于“老”的记录都跟着“丁”或是“退”等，指代那些年老的男性（老丁）或者年老退休的（老退）的个人。“老”意味着某个人曾经作为特殊义务人群但当年纪达到一定岁数后不再负有特殊义务。

和 RETIRED 退休一样，那些标识有“老”的记录中，个人的年纪也有较宽的范围，当然也有一定模式。如表 10 中列出的，约 86%的被标记为“老”的个人都是在其年纪超过 60 岁后才首次被记录为年老个人的。同时，却也有个别个人年纪小于十岁而被标注为“老”。Lee and Campbell (1997,165-66)讨论了那些被记录为年老和退休的人群的年龄情况。

表 10 个人第一次被标注为“老”时的年龄分布

年龄	频数	百分比	累计百分比
缺失值	448	6.08	6.08
1-10	2	0.03	6.11
11-20	5	0.07	6.18
21-30	19	0.26	6.44
31-40	27	0.37	6.81
41-50	44	0.60	7.40
51-60	505	6.86	14.26
61-70	3,868	52.53	66.78
71+	2,446	33.22	100.00

HAS_POSITION 在职

HAS_POSITION 是表示一位男性在进行人口登记的时候拥有带俸禄的职位的标识变量。CMGPD_LN 中主要有三类人群有俸禄：官员、大部分的军人以及全部的工匠。俸禄主要有两种形式：钱和粮食。因为带俸禄的职位通常只由男性担当，故对所有的女性这个变量的值设为缺省（-98）。

鉴于 CMGPD_LN 中所涉及的人群主要为农业人口，所以拥有一个带俸禄的职位不仅意

味着个人地位上的成就，也意味着额外的收入和在家中的特权。如表 11 所示，在 CMGPD_LN 中所有人口中，只有 2.2%的 18-60 岁男性拥有带俸禄的职位。

表 11 拥有带俸禄职位的 18-60 岁男性分布

带俸禄的职位	频数	百分比
0	509,914	97.8
1	11,403	2.2
总计	521,317	100

HAS_POSITION（在职）这里所涉及的地位高低的范围是很宽的，从最基本的带薪职位比如工匠和士兵——分别用 ARTISAN 工匠和 SOLDIER 士兵变量加以区分，到权力更大，俸禄更为优厚的高等职位。在我们的分析版本中不仅提供了这些原始的职位信息，还增加了额外的变量来方便对个人的年收入进行估计。

因为 HAS _POSITION 是通过对原始户口册中的个人的行政地位进行文字分析得到的，故 HAS _POSITION 变量自身并不能说明在进行登记的时候该职位是否依然有效。某位男性曾经有过职位但已经退休，并在户口册中记录为以该职位退休，其 HAS _POSITION 变量和 RETIRED 变量值都为 1。这一情况对于分析高龄的男性不构成问题，因为一般我们关心的就是某个个人在其年轻的时候是否有过职位。尽管如此，如果要分析有效的带俸禄职位的情况，使用者可以用下面的代码：

```
HAS_POSITION==1 & RETIRED!=1 & OLD!=1
```

ARTISAN 工匠

ARTISAN 工匠也是一个表示个人拥有工匠职位的标识变量。工匠的地位主要是通过个人的“匠”字称谓来确定，类似于英语中的 smith 的后缀。例如，carpenters 是木匠，dyers 是染匠等等。CMGPD_LN 中所涉及的人口包含契约劳力，所以工匠的职位不仅意味这些人拥有特定的技能，还说明他们需要使用技能来为皇家内务府做工，而获得相应报酬。在十九世纪晚期，一般一个工匠的年收入约为 6 两白银。[30]

如表 12 所示，在观察期间，只有 0.4%的 18-60 岁男性曾为工匠。

表 12 曾为工匠的 18-60 岁男性分布

工匠	频数	百分比
0	519,398	99.6
1	1,919	0.4
总计	521,317	100.0

因为程序主要是通过对原始户口册中的个人行政地位进行文字分析得到ARTISAN变量的值，故 ARTISAN 变量自身并不能说明在进行登记的时候该职位是否依然有效。若需确定

[30] 见 Guochao Jianye Chuji Jilue 1983 年版，卷 5，616。我们希望感谢 Mark Elliott 让我们注意到了这份盛京内务府的宝贵的资料。

有效的工匠职位的情况，使用者可用下面的代码：

```
ARTISAN==1 & RETIRED!=1 & OLD!=1
```

SOLDIER 士兵

SOLDIER 是显示某个人拥有士兵地位的标识变量。士兵地位的判断主要依据原始户口册中，个人具体的行政地位信息中出现“兵”或“甲”的字样。如表 13 中显示，在我们所涵盖的整个时期中，只有 1.41%的 18-60 岁男性曾经有过士兵的职位。拥有士兵的职位不仅意味着相应的收入还包括一定的特权。在十九世纪晚期，一名士兵的年收入有 24 量白银，远远超过工匠的收入[31]。

表 13 曾为士兵的 18-60 岁男性分布

士兵	频数	百分比	
0	513,963		98.59
1	7,354		1.41
总计	521,317		100.00

和其他描述行政地位的变量一样，SOLDIER 变量也是通过对原始户口册中行政地位信息进行解析后得到的，其本身无法确定该职位在进行登记时是否依然有效。若需确定士兵的身份，使用者可以用下面的代码：

```
SOLDIER==1 & RETIRED!=1 & OLD!=1
```

NO _STATUS 无职位

NO _STATUS 无职位也作为一个标识变量来显示某个人在户口册中没有记录其任何行政地位。值 1 表示没有行政职位，0 表示有。大部分男性都有其行政地位，但并非全部。那些年纪较小还不能从事劳役服务的男性一般都没有行政地位。鉴于女性是不能拥有这类职位的，所以女性的 NO _STATUS 的值设为缺省（-98）。

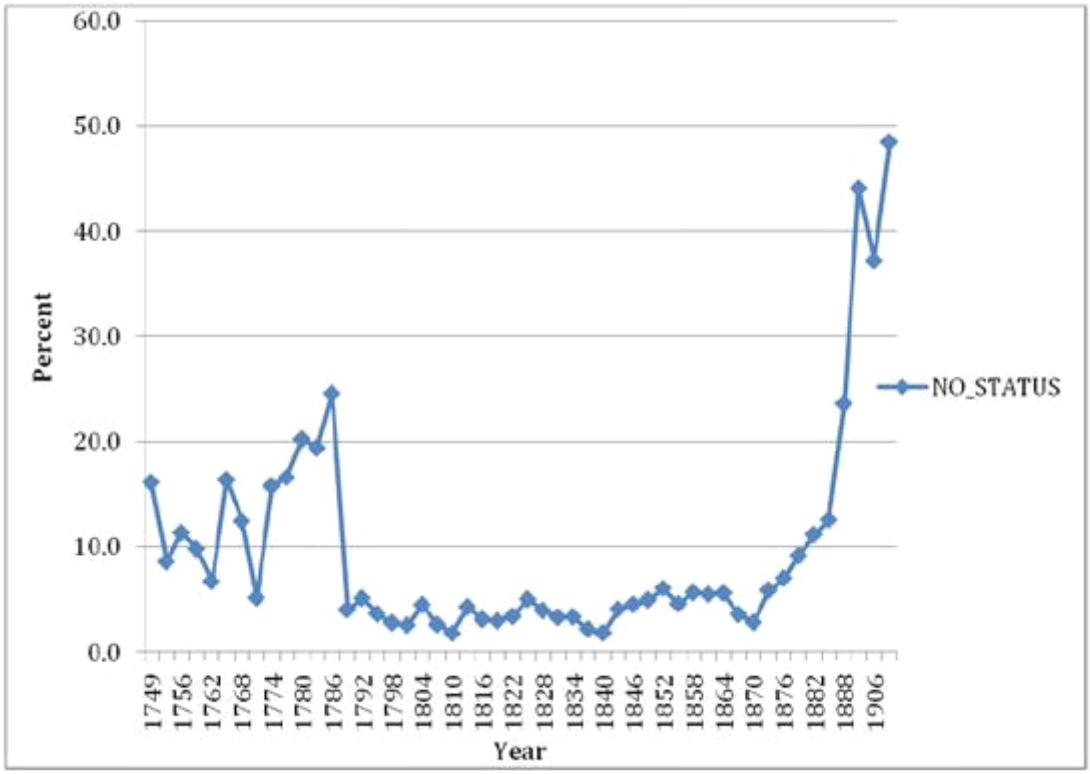
男性一般在 18 岁成年后才会拥有其行政地位，故这其中的身份转变是突然的。17 岁男性中有约 86.5%的个人没有行政地位，而 18 岁男性中，却只有 38.2%依然没有行政地位。随着年纪的增长，到 30 岁约只有 10%的男性没有行政地位，这些男性多半是因为身体有残疾或比较虚弱。总体来看，如表 14 所示，CMGPD_LN 中有约 12%的男性是没有其行政地位的。

表 14 18-60 岁男性无职位分布

NO_STATUS	频数	百分比	
0	457,352		87.73
1	63,965		12.27
总计	521,317		100.00

[31] 见 Guochao Jianye Chuji Jilue 1983 年版，卷 5，615。

在 CMGPD_LN 中，行政地位头衔的分布在不同时期和不同人口群体的数据库中均有不同。图 9 中显示的即在 1749-1789 年间，男性中没有行政地位的个人所占比例在 4%到 25%间浮动。当然这其中的浮动有一部分是源于某些年代有效户口册的稀少。我们还发现，有一些人群中没有行政地位的男性成员更多些，所以当碰到某一年代，恰好这类人群的户口册留存多，比较普遍，则该年代整个人群中没有行政地位的男性的比例就会上升。1789 年至 1870 年间，这一比例一直维持在 5%左右，然而自 1873 年后，该数字直线上升，到 1909 年达到约 50%。[32]



注：Year：年份 Percent：百分比

图 9 18-60 岁男性无行政地位的比例，1749-1909

ZU ZHANG 族长

ZU ZHUANG 族长是显示某个人作为一族首领的虚拟变量。在八旗制度中，关于于“族”的组织设置源于女真部落时期。随着满洲国和清代对其进一步的发展，这一“族”的设置演化成高度制度化的模式。朝廷推出并逐渐完善了族长的制度，任命一族的首领作为管理其下旗人的朝廷代理人。他们的职责包括主管婚姻、遗产分配、葬礼、祭祀等族中的重要事件。

在辽东八旗户口册中，“族”及“族长”的概念出现在十八世纪八十年代。1789 年前，户口册中用来组织人口的基本单位是“一户”。这里“一户”即指一族，每“一户”都有其首领，也称为族长，其属下的所有家族成员都根据其与族长的关系来列在族长之后。自 1789 年起，人口登记中把原来的“一户”分成了多个户，记为“另户”。因此，“族”便作为一个

[32] 在他们 2004 年对辽东地区八旗人口的研究中，Yizhuang Ding, Songyi Guo, James Lee, 和 Cameron Campbell 给出了不同时间行政地位头衔不同的几个具体例子 (2004, 93-98)。

正式的人口单位来管理其下的所有户。对正身旗人和特殊义务旗人人口来说，族长拥有行政管理的权威，因此其中的每个个人都依等级制从属于由高到低（佐领，族，户）的某一阶层。

族长的头衔会独立于个人的行政地位信息而单独记录，因此 ZU_ZHANG 是作为一个独立的变量来描述个人的行政地位。而且“族长”并不是一个带俸禄的头衔，也不是一个永久的地位，一族有时会改变其族长的人选。

3.C.ii 身份识别变量

数据库中所有的身份识别变量都是由软件生成的。根据功能可以将其划分成两类：便于进行数据管理和分组而设立的；为进行数据关联而设立的。像 RECORD_NUMBER（记录编号）、REGISTER_SEQ（户口册序号）、ZU_SEQ（族序号）、HOUSEHOLD_SEQ（户序号）等就属于第一类；而像 PERSON_ID（个人身份编号）、 MOTHER_ID（母亲身份编号）、FATHER_ID（父亲身份编号）、FATHER_ID_IMPUTED（父亲身份假定编号）、GRANDFATHER_ID（祖父身份编号）、GRANDFATHER_ID_IMPUTED（祖父身份假定编号）、WIFE_1_ID（妻子 1 身份编号）、WIFE_2_ID（妻子 2 身份编号）和 HUSBAND_ID（丈夫身份编号）则属于第二类。关联变量是 CMGPD_LN 的一大特色，其主要通过编码员对不同户口册进行关联以及由 RELATIONSHIP 变量所决定的亲属间的关系来生成，这些关联变量把同一个人在不同户口册中的记录事件建立关联（PERSON_ID），同时也把个人和其他家庭成员建立关联（MOTHER_ID）。

这里需要提醒的一点是，鉴于所有的身份识别变量都是通过复杂的计算机程序对 RELATIONSHIP 信息进行分析加上编码人员的建立关联来生成的，所以并不一定是完美无缺的。因为大部分涉及的关联关系都是相对直接的，所以如果原始户口册中出现的错误或是编码者在录入过程中出现错误都可能造成亲属间错误的关联关系。对于一些无法确认同某亲属关系的情况，程序会采用简洁的办法获取。比如，某个人的父亲无法确认，该情形出现可能是因为某个人在其第一次出现在户口册之前父亲便已过世，这种情况下程序便会查找该个人的兄弟姊妹中是否有记录的父亲身份，若信息有效就把该信息复制过来。当然，很有可能，这种推测本身就是错误的。我们在建立亲属间关联时所用到的全部信息即是基本版中的 RELATION_SHIP、MARITAL、STATUS、AGE_IN_SUI 和 HOUSEHOLD_ID 等变量，使用者可以根据自身的技术来编写自己的程序来建立这类关联。

RECORD NUMBER （记录编号）

是一个顺序编号来确定 CMGPD_LN 中的某一记录在原始户口册中的位置，它是根据 DATASET、YEAR、REGISTER_SEQ 变量进行排序的，其值唯一。生成该变量的代码如下：generate record_number =string(_n, "%09.0f")。每一编号唯一对应整个数据库中的某一记录。

这一变量是作为用户将 CMGPD_LN 基本版中变量和 CMGPD_LN 限制分析版中变量结合的关键关联变量。也可以作为用户常用的将带有新生成变量的数据库和基本数据库结合的记录标识。当然，如果用户对数据库按照某一变量进行过排序的话，RECORD_NUMBER 还可以用来恢复其原始的顺序。

作为使用RECORD_NUMBER的一个例子,下面的程序载入了基本版数据，并合并进了分析版数据。

```
use"..\CMGPD-LN from ICPSR\ICPSR_27063\DS0001\27063-0001-Data.dta"
merge 1:1 RECORD_NUMBER using"..\CMGPD-LN from
ICPSR\ICPSR_27063\DS0003\27063-0003-Data.dta", nogenerate
```

这个程序的假设是从 ICPSR 下载的包含数据库的压缩文件被放在邻近的文件夹中，解压缩，所以.dta 文件会在子目录 ICPSR_27063\DS0003 和 ICPSR_27063\DS0001 中。

PERSON ID 个人身份编号

PERSON ID 是标识不同个人的变量。每一编号唯一确认数据库中某一个人的全部记录。该变量通过两个阶段来构造：第一阶段中，编码员人工的把户口册中个人同上一次户口册中属于他的记录建立联系，通过指定该记录的编号实现。因为在连续的两次户口册中，都是根据同样的顺序记录户和其成员，所以方便编码人员通过比较个人的姓名以及上下文的信息，加上居住村庄的信息就可以确定不同户口册中的同一个人。在第二阶段，程序把每一次户口册中的信息与上一次结合，最终生成某个人在不同户口册中的所有相关记录的标识。

PERSON_ID 变量可以帮助用户按个人对记录进行分组。结合 YEAR 变量，用户很方便就可以把记录按照个人每一年的顺序进行安排。在 STATA 程序中使用代码为：sort person_id year。在 by person_id year（按个人年份安排）基础上，用户可以对某个人一生的所有记录进行各种操作。比如，在按照 PERSON_ID 和 YEAR 排序后，by person_id year: generate birthyear_first =birthyear[1]就可以创造一个新变量来显示个人的出生年月，继而可以计算出他一生中自第一次出现在户口册中开始，每一条记录发生时对应的年纪。

MOTHER_ID 母亲身份编号

MOTHER_ID 提供了所有男性和未婚女性母亲的 PERSON_ID 信息，及所有已婚女性和寡妇的婆婆的 PERSON_ID 信息。该变量也是软件通过在户口册中寻找个人和其母亲的关系来确定。举例说明，当程序发现某 1s1s 的个人，就会在该户中寻找 1sw 对应的个人。程序对同一个人在其出现的全部的户口册中进行查询，当不同户口册信息间出现不一致时，按照最早的户口册中显示的关联信息来确定。当程序通过直接的方法无法确定关系时，它就通过间接的方法来查找：比如，查询个人是否存在其父亲有一个妻子在个人还未出现在户口册之前便已过世的情况，亦或者个人的兄弟姊妹中其他人有确定的母亲。最后对于那些软件实在无法确定的个人，该变量的值设为-99。对母亲连接变量进一步的讨论包括一些连接过程中关键假设的回顾，见关于 M_ID_1, M_ID_2, M_ID_3, and M_ID_4 的讨论。

这一变量可以方便用户按照母亲对个人进行分组，还可以链接到个人母亲的相关记录来收集关于她的信息。

FATHER_ID 父亲身份编号

FATHER_ID 提供了所有男性和未婚女性的父亲的 PERSON_ID 信息，以及已婚女性和寡妇的公公的 PERSON_ID 信息。FATHER_ID 的值也是程序通过两种方式获得的：对于个人的父亲在原始记录中明确的，该值便是其父亲的 PERSON_ID，即通过原始记录中的人际关系来建立联系；对于个人的父亲信息没有出现在原始数据中的，该值假定。

这一变量方便用户对个人按照其父亲的信息来分组和查询其父亲的相关信息。

FATHER_ID_IMPUTED 父亲身份假定编号

FATHER_ID_IMPUTED 作为一个标识变量来标明这一记录中的 FATHER_ID 所对应的个人是否是数据库中所包含的个人。建立 FATHER_ID 变量是为了方便对那些根据 RELATIONSHIP 可以确定为兄弟姊妹的个人进行分类，包括那些父亲不在数据库范围中的个人——一般是因为最早期户口册中出现时，他们的父亲已经过世而没有出现在数据库中。举个例子，如果在最早期户口册中显示有兄弟姐妹几个人生活在一起，父亲已经过世，这时我们给他们共同的父亲一个假定的身份编号，就使得在分析的时候，可以将他的子女分组到一起，无论父亲是否真的在户口册中出现过。如果 FATHER_ID_IMPUTED 值为 1，此时的 FATHER_ID 可以用来对其后代的记录进行分组；但并不能用于搜集父亲的相关信息。

GRANDFATHER ID 祖父身份编号

GRANDFATHER_ID 提供了男性和未婚女性的的祖父的 PERSON_ID 以及已婚女性和寡妇的丈夫的祖父的 PERSON_ID 信息。像 FATHER_ID 变量一样，对那些已婚的和寡妇的个人，GRANDFATHER_ID 值也是由软件生成，是其丈夫的祖父信息。对于个人的祖父信息在原始记录中可以通过搜索 RELATIONSHIP 变量而明确的，其值为祖父的原始 PERSON_ID 信息。举个例子，某头领的孙子，1s1s，就可以直接关联到他的祖父——头领, e。当个人的祖父信息无法直接从原始数据得到——大部分情形是祖父已过世——一般认为祖父是个人父亲的父亲，故将个人父亲对应的 FATHER_ID 信息直接拷贝到该个人的 GRANDFATHER_ID 信息中。

GRANDFATHER ID IMPUTED（祖父身份假定编号）

GRANDFATHER_ID_IMPUTED 也是一个用来显示某一记录中的 GRANDFATHER_ID 变量是否对应于一个未包括在原始数据中的个人，并且设定一个标识符来帮助对其孙辈进行分组。具体可以参见 GRANDFATHER_ID_IMPUTED 变量的注释。如果 GRANDFATHER_ID_IMPUTED 变量值为 1，则 GRANDFATHER_ID 只能用于对其孙辈个人的记录进行分组。

WIFE_1_ID 第一个妻子的身份信息

WIFE_1_ID 显示的是某个人在当前户口册中的妻子的 PERSON_ID。程序是依据当前的原始户口册中 RELATIONSHIP 信息建立个人和其妻子的关联而生成的该变量。对与所有在 RELATIONSHIP 变量末尾有“w”的女性，程序便搜索这户人中的男性中有一样的 RELATIONSHIP 信息但末尾没有“w”的个人作为其丈夫。举例说明，某户中一女性 RELATIONSHIP 值为 2ybw——排行第二的弟兄的妻子，程序便搜索 RELATIONSHIP 值为 2yb 的男性。对于户主的妻子，RELATIONSHIP 值为 w，程序则寻找 e。对于户主的母亲，m，程序搜索父亲，f。当程序搜索到对应个人后，就将女性的 PERSON_ID 值拷贝到男性的 WIFE_1_ID 中，将男性的 PERSON_ID 值拷贝到女性的 HUSBAND_ID 中。

但鉴于一夫多妻和丧夫女性再嫁的可能性，我们的程序并不打算去裁决对于同一个出现在不同户口册或是同一户口册中的多个妻子之间的身份。如果在同一户口册中，某个人有两个相关联的妻子，则把第二个妻子的 PERSON_ID 信息存储到 WIFE_2_ID 变量中。如果对

于某个人在户口册中显示的妻子同前次登记不同，那么很自然的，WIFE_1_ID 的信息也不同。

因为 WIFE_1_ID 和 WIFE_2_ID 的值都是按照当前户口册中的内容确定的。故用户若要搜集某一鳏夫已去世妻子的信息，我们还需要增加额外的处理，即把在最初其妻子仍被记录时的 WIFE_1_ID 信息向后拷贝。

女性、鳏夫和未婚男性的 WIFE_1_ID 和 WIFE_2_ID 值设为缺省（-98）。

WIFE_2_ID 第二个妻子的身份信息

WIFE_2_ID 提供的是在原始数据中显示的个人的第二个妻子的 PERSON_ID 信息。和 WIFE_1_ID 一样，它的值也是由专门的计算机程序根据原始记录将妻子的信息关联到丈夫上。从根本上说，如果超过一位以上的女性的 RELATIONSHIP 信息显示其作为妻子，那么将其的 PERSON_ID 的信息放进该男性的 WIFE_2_ID 中。由于在我们现有的户口册中所记载的人群中一夫多妻现象极其稀少，所以这类情形很少见。

HUSBAND_ID 丈夫身份信息

HUSBAND_ID 记录了在当前户口册中，某位女性的丈夫的 PERSON_ID 信息。它是作为根据 RELATIONSHIP 信息确定 WIFE_1_ID 和 WIFE_2_ID 的一部分来同时确定的。当然情形是一样的，该变量只指代当前户口册中显示的丈夫。

MARTIAL_STATUS 婚姻状态显示为未婚女性和丧夫的女性 HUSBAND_ID 值设为缺省（-98）。同样，所有男性的该值也为缺省。另外，对于一些 MARTIAL_STATUS 婚姻状态显示为已婚，但其丈夫却不在数据库范围内，这种情况大部分是因为她们是在上一次户口册中显示已经外嫁，故其婚姻状态为已婚，她们的 HUSBAND_ID 设为-99。

一共有 138 名女性在不同户口册中与不同丈夫相连。我们还没有调查这些案例。因为 RELATIONSHIP 是 HUSBAND_ID 的基础，这些案例可能反映了 RELATIONSHIP 编码过程中的不一致。

REGISTER_SEQ 登记序列号

REGISTER_SEQ 是一串指示户口册中某一具体年份的标识符。它的值是由录入者根据原始资料确定的，和 DATASET 和 YEAR 变量在一起，但其自身也能唯一确定某记录。依据 DATASET、YEAR 和 REGISTER_SEQ 的值对记录进行排序就可以使数据恢复到最初顺序。

ZU_SEQ 族序列号

ZU_SEQ 是一串确定相关记录属于某一特定族特定年份的标识序列号。其值也是由编码者对原始数据的录入确定的。在原始户口册中，列的顶端的“一户”字样即说明这是新开始的一族成员。每一次的户口册中，最先出现的族的族序列号为 1，其他族依次按顺序排列。

HOUSEHOLD_SEQ 户序列号

HOUSEHOLD_SEQ 是用来确定某一特定族中的不同户的标识符。每一族中的第一户的值为 1.考虑只在 1789 年以后才开始以户为单位，故在 1789 年前该值都设为缺省。它的值同样是由编码人员依据原始数据录入的。自 1789 年以后，很容易就可以在原始户口册中区分不同户，即每一户的头一列中都包括有“另户”的注释，而在一户中最先列出的即是该户的户主。

HOUSEHOLD_ID 户编号

HOUSEHOLD_ID 唯一的确认属于某一户成员某一年份的所有记录。这一变量也是由计算机软件生成，和 DATASET、YEAR、ZU_SEQ 和 HOUSEHOLD_SEQ 一起。创立这一变量的主要目的便是分组，举个例子，需要把相关记录按照户进行分组，再创立一个新变量来显示相关记录的条目数量，就可以用下面的 STATA 代码：

```
sort household_id
by household_id: generate household_observations = _N
```

3.C.iii 分析变量

在反映户口册中基本人口信息的变量和身份鉴别变量之外，我们还以此为基础创建了一系列分析变量来便于用户做人口分析时使用。鉴于在分析版中提供了诸多这类变量，在此版本中我们主要介绍一些基本的分析变量，包括“和个人一起居住的近亲数量”“某一事件发生的情况”“在特定类型的分析中，鉴别某些记录是否有效的被包含在分析中”等等。

3.C.iii.a 计数变量

BROTHER_COUNT 兄弟个数

BROTHER_COUNT 兄弟数目统计了当前户口册中，该户中某个人的兄弟数目。它的值是由计算机程序根据查询和该个人有相同 FATHER_ID 的男性成员的数目来得到的。对 1789 年之前的数据，所有 BROTHER_COUNT 的值为缺省（-98），因为在那以前，户口册中并没有按户进行划分。

SISTER_COUNT 姐妹个数

SISTER_COUNT 姐妹数目统计了当前户口册中，该户中的某个人的姐妹数目。其值也是由计算机程序根据查询一户中和该个人有相同的 FATHER_ID 的未婚女性数目得到的。同样的，1789 年以前该变量设为缺省（-98），因为那时无法分辨同属一户的成员。

MALE_COUSIN_COUNT 堂兄弟个数

MALE_COUSIN_COUNT 堂兄弟数目是统计当前户口册中某个人的父堂表兄弟数目的变量。其值是计算机程序查询一户中与该个人有相同 GRANDFATHER_ID 值的男性成员得到的。对于 1789 年之前，该变量设为缺省（-98）。

FEMALE_COUSIN_COUNT 堂姐妹个数

FEMALE_COUSIN_COUNT 堂姐妹数目是统计当前户口册中某个人的父系堂姐妹数目的变量。其值是计算机程序查询一户中与该个人有相同 GRANDFATHER_ID 值的其他未婚女性成员得到的。对于 1789 年之前，该变量也被设为缺省（-98）。

UNCLE_COUNT 叔伯个数

UNCLE_COUNT 叔伯数目是统计当前户口册中一户中某个人父亲的兄弟数目的变量。其值是计算机程序查询该户中是否有男性成员的 FATHER_ID 值与某个人的 GRANDFATHER_ID 值相同，并统计其数目。对 1789 年之前记录，该变量被设为缺省（-98）。

AUNT_COUNT 姑姑个数

AUNT_COUNT 姑姑数目是统计当前户口册中一户中某个人父亲的姐妹数目的变量。其值是计算机程序查询该户中是否有未婚女性成员的 FATHER_ID 值与某个人的 GRANDFATHER_ID 值相同，并统计其数目。对 1789 年之前记录，该变量被设为缺省（-98）。

FATHER_ALIVE 父亲在世

FATHER_ALIVE 是一标识变量。其值为 1 表示某个人的父亲在当前户口册中仍健在。值为 0 则表示其父亲当前已经过世，也可能是没有在该户的成员中。鉴于极少出现父亲仍在世，儿子却已离开家的情形，因此第二类情形在 CMGPD_LN 中是个别特例。

MOTHER_ALIVE 母亲在世

MOTHER_ALIVE 母亲在世是一标识变量。其值为 1 表示某个人的母亲在当前户口册中仍健在。值为 0 则表示其母亲当前已经过世，也可能是没有在该户的成员中。然而，和父亲的情形类似的，儿子极少和母亲生活在不同户中。母亲在世时，儿子一般极少脱离户成员；而寡妇在她仍有儿子活着时，也极少改嫁。

SON_COUNT 儿子数目

SON_COUNT 儿子数目反映的是某一个人至当前登记时，所有生存到已被记录时儿子的数目。其值主要是程序依据记录之间的关联查询出的。它不是依据在当前户口册中个人的儿子数目，而是分别计算当前户口册中一户里男性成员的 FATHER_ID 或 MOTHER_ID 值是某个人的 PERSON_ID 的数目，以及根据年龄计算在当前登记时已经出生的儿子的数目。SON_COUNT 的值可能与当前登记中儿子的数目不一致，因为会包括那些出现在后来户口册中，而实际上在当前登记时已经出生的儿子。这一变量是考虑到对个人的生育细节进行分析时的便利而设立的。当然，这一变量很可能也是低估了个人实际的出生的儿子数量，如前文中所分析过的，许多孩子在婴儿期或是幼年时便夭折而根本没有出现在户口册中。因此，SON_COUNT 应该是作为个人所有至少活到被登记时儿子的数目。

DAUGHTER_COUNT 女儿数目

DAUGHTER_COUNT 女儿数目统计了个人所有在当前登记时所有女儿的数目。具体可以参见 SON_COUNT 变量中的讨论。考虑到登记中，许多女儿都没有被实际记录，所以此

变量无疑大大的低估了实际数目。不同年代、数据库、居住地中对于女儿的登记情况都有所不同，所以用户在使用此变量做分析之前，须检查相关的描述统计量来熟悉和辨别其所属的类型。

3.C.iii.b 事件发生的标识变量

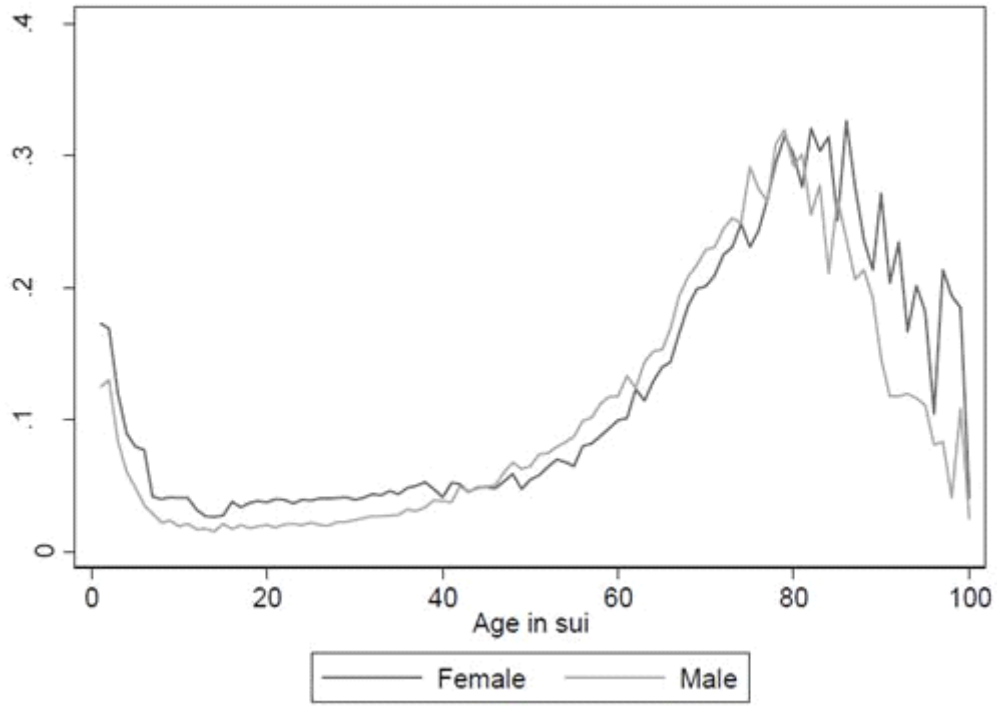
这里所有的标识变量都是用来表示某个人在下次户口册中将要发生某类人事事件。创建这些变量同样也是为了便于对人口学事件进行分析。它们在事件历史分析中尤其有用——即把个人在未来将要发生的事件和当前或是过去该个人的某些特质相联系。

对于每个人口群里最后一本户口册，在当前户口册与下次户口册之间的事件发生的标识变量被设置为结构性缺失 (-98)。但是我们强烈建议用户使用已创建的标识变量例如 AT_RISK_DIE 来分析死亡率和结婚率，以便在研究中对观察值做出适当的取舍。我们强烈鼓励用户使用附加标识变量如 NEXT_3 和 NEXT_6，他们代表下一个有效观察值是否在 3 年或者 6 年之后出现，以便将研究限定于连续时间段内的当前或下次户口册的观察值。

NEXT DIE 即将去世

NEXT_DIE 是一标识变量。值 1 表示某一个人在下次可查的户口册中被注释在该户口册相关的过去三年间去世。它的值是由程序根据 DIED 变量的情况得到的。此变量的主要作用是进行死亡情况分析。鉴于此变量是依据原始户口册中的注释设定，它只能用于 CMGPD_LN 中那些存在下一次三年期户口册的情形。倘若在当前户口册和下一次有效的户口册之间有过多次登记，该变量并不能包含那些出现在已经缺失的户口册上的死亡注释，而只包括那些出现在下一次有效的户口册中记录的死亡事件。 AT_RISK_DIED 变量涉及就是当前存在而下一次三年期的户口册也存在的情况。对于在当前登记中被记录为自上次登记后已死亡，即个人的 PRESENT 变量值为 0 的个人，NEXT_DIE 值设为缺省（-98）。

在研究老龄人口死亡率时应当十分注意，因为有些可能已经死亡的人口没有被记录死亡，使得观察值数量比实际数量要大。一小部分人没有任何死亡记录，却被一册册地记录下来。他们中的许多在生命中的某一刻被记录为 ABSCONDED。因为他们仅仅代表了一小部分人，所以这些观察值的存在在幼龄时期不重要，但是当人们去世越来越少的人存活时，他们代表了越来越大的人口份额，使得计算死亡率的分母虚大。之前利用 CMGPD-LN 进行的死亡率的研究,例如 Lee 和 Campbell (1997)的研究,都把样本选定在 75 岁以下。图 10 代表了未来三年不同年龄的死亡概率，其中老龄阶段人口死亡记录的问题很明显。特定年龄的死亡概率直到 80 之前一直上升，随后下降。图 10 也显示了在 40 岁之前，女性死亡率比男性死亡率高，40 岁之后比男性死亡率低。



注：Female：女性 Male：男性

图 10 男女不同年龄未来三年的死亡概率

NEXT_MARRY 即将婚嫁

NEXT_MARRY 也作为标识变量。值 1 表示该个人在下一次户口册中被记录已结婚。对于男性，程序根据下一次登记中出现其妻子而确定，表示个人是否在此次登记和下次登记之间结婚。若在分析男性婚姻情况，需要把那些没有包括在 CMGPD_LN 中的缺失户口册考虑在内的话，用户即需要考虑计算此次登记和下次有效登记之间的情况。该变量没有包括某些男性的妻子在下一次有效登记之前便去世的情形。AT_RISK_MARRY 变量显示的就是那些当前未婚，且下一次三年期户口册存在的男性。

对于女性，NEXT_MARRY 是程序根据在下一次有效登记中显示该女性已外嫁的来确定，正如 MARRY_OUT 变量显示的一样。如果某个人在显示其外嫁户口册所涉及的三年内外嫁，该变量就会被设为 1.类似的，NEXT_MARRY 的使用也会受到核 NEXT_DIE 一样的限制，所以建议用户使用 AT_RISK_MARRY 来把分析对象限制在那些当前未婚且下一次三年期登记存在的情形下。

对于那些被注明已经离开的个人，即 PRESENT 变量值为 0 的个人，NEXT_MARRY 值为缺省（-98）。

NEXT_REMARRY 即将再婚

NEXT_REMARRY 是一标识变量显示在下一次登记中某位丧偶的个人再婚事件。对于丧妻的男性而言，该值是程序通过查询当前记录为丧偶的男性个人在下一次登记中出现其妻

子的事件而确定的。再婚的时间可能是当前登记和下次登记之间的任意时间。考虑到和NEXT_MARRY 同样的局限性，我们设立 AT_RISK_REMARRY 变量来鉴别那些当前显示为丧偶的男性，且其下一次三年期的登记为有效的情况。

对于女性而言，NEXT_REMARRY 变量的值则是程序依据下一次登记中显示的再次外嫁记录而确定，即由 REMARRY_OUT 变量所显示的情况。因此，只有对于那些在下一次登记所涉及的三年间再次嫁出的女性个人，其 NEXT_REMARRY 变量值为 1。用户在使用此变量时同样也需要借助 AT_RISK_REMARRY 来限制其应用的范围，具体可以参照 NEXT_MARRY 和 NEXT_DIE。

对于在当前登记中显示已离开的个人，即其 PRESENT 值为 0 的情形，他们的 NEXT_REMARRY 变量值设为缺省-98。

NEXT_ABSCONDED 即将逃

NEXT_ABSCONDED 是一标识变量，值 1 表示某个人在下一次登记中被记录在过去三年间潜逃离开。该变量是程序依据下一次户口册中的 ABSCONDED 变量的值来确定的。鉴于很多个人在其案件未结束前，会反复被记录为潜逃情况，所以他们会多次出现 NEXT_ABSCONDED 值为 1 的情况。所有女性的该变量值设为缺省（-98），因为潜逃事件并不适用于女性。

NEXT_BOYS 即将有男孩出生

NEXT_BOYS 变量记录了某一个人在当前登记与下次登记之间拥有新降男孩儿的数目。其值是程序依据记录间关联来确定的，具体来讲，即通过计数那些拥有 FATHER_ID 或 MOTHER_ID 指向我们所考虑个人的男孩子数目，且他们的出生年月根据推算恰好在两次登记之间。此变量反映的情况可能和人为的把两次户口册中显示的儿子数目进行比较的结果不同，因为可能存在某一实际已出生的孩子在当次户口册中未被记录，而在之后出现的情形。当然，这一变量很可能也是低估了实际的数量，如前面讨论过的，许多在婴儿期或幼年期夭折的孩子并不出现在户口册中。准确的说，此变量记录了某一个人在当前登记与下一次登记之间新出生的，并活到被记录时的男孩数目。

NEXT_GIRLS 即将出生女孩

NEXT_GIRLS 计数了某个人在当前户口册和下次有效户口册之间新出生的女孩数目。具体可以参照 NEXT_BOYS 变量的生成方法。鉴于原始记录中很大部分的女孩出生记录都被省略了，所以此变量无疑也是大大低估了实际情形。用户须在检查相关的描述性统计结果，了解清楚具体的世代、地点和数据库后再使用此变量进行分析。

NEXT_3 下个三年

NEXT_3 是一虚拟变量。值 1 表示某一个人下一三年期的相关记录存在于数据库中，其值由程序自动生成。设立该变量是便于进行事件历史分析，即确定与哪些记录相关的下个三年期的有效登记存在，用户就可以解决因缺失某些户口册而带来的问题。

NEXT_6 下个六年

NEXT_6 也是一虚拟变量，和 NEXT_3 类似。值 1 表示某一个人下一次有效的登记是在 6 年后。

3.C.iii.c 确定相关事件的潜在人群的标识变量

下面介绍的这些变量是为了确定相关人事事件的潜在人群而创建的。它们在进行事件历史分析时尤其有用。

AT_RISK_DIE 死亡风险

AT_RISK_DIE 作为一个标识变量，通过其可以鉴别出那些适用于死亡分析中的记录。其值是依据 PRESENT 和 NEXT_3 变量共同决定的。当 PRESENT 和 NEXT_3 变量值均为 1，即某一个人在当前登记中显示为存活且下次三年期登记有效，那么他的 AT_RISK_DIE 值为 1，其他情形下为 0。对注释自上次登记后已经离开的个人，该值设为缺省（-98）。

AT_RISK_MARRY 结婚风险

AI_RISK_MARRY 作为一个标识变量，其帮助鉴别那些适用于进行首次婚姻分析的记录。它的值由程序依据 PRESENT、NEXT_3 以及 MARTIAL_STAUS 来决定。当 PRESENT 和 NEXT_3 值为 1，MARTIAL_STAUS 显示为已婚，即说明某个人在当前登记中存在并未婚，而下一次三年期的登记有效，则把其 AT_RISK_MARRY 变量值设为 1，其余情况值为 0。对登记中显示自上次登记后已离开的个人，该变量设为缺省（-98）。

AT_RISK_REMARRY 再婚风险

AT_RISK_REMARRY 作为一个伪变量来鉴别那些适用于再婚分析的记录。它的值由程序依据 PRESENT、NEXT_3 以及 MARTIAL_STAUS 来决定。当 PRESENT 和 NEXT_3 值为 1，MARTIAL_STAUS 显示为丧偶，即说明某个人在当前登记中存在并显示为丧偶，而下一次三年期的登记有效，则把其 AT_RISK_REMARRY 变量值设为 1，其余情况值为 0。

EVER_ABSCONDED 曾经出逃

EVER_ABSCONDED 作为一个标识变量，通过其可以鉴别出个人在现在、过去或者是未来的记录中是否有过 ABSCONDED 的记录。我们强烈建议使用者通过 NEXT_DIE 或者其他排除了被标识了 EVER_ABSCONDED 的人口的变量来检验死亡率。因为在我们自己的研究中发现，即使现在的记录中没有被列为 ABSCONDED，但在过去的记录中有过 ABSCONDED 的人口，一般死亡率比较低。这些人代表了一大部分异常长寿的人群。

3.D分析版数据（DS3）

分析版数据主要包括计算机程序所产生的标识变量和分类变量。这些变量主要基于个人的登记姓名、身份地位和其它特征，或者家庭中成年男性的登记姓名和其它特征而创造的。其中许多都是表示个人官方职位和头衔的标识变量，根据职位和头衔的汉语拼音构造而成。

将分析版数据和基本版数据结合使用，需要利用 RECORD_NUMBER 来合并两个版的数据。每个 RECORD_NUMBER 能够在整个数据集中标识出一个特定的观察值。RECORD_NUMBER 同时出现在基本版数据，分析版数据，血亲版数据和限制版数据中，相同的观察值在不同数据集中的 RECORD_NUMBER 相同。因此，我们可以利用 RECORD_NUMBER 把限制版数据和分析版数据中的信息与基本版数据中具有相同 RECORD_NUMBER 的特定观察值连接起来。如果一个数据集已经加载到内存中，以下 STATA 代码可以将它合并进另一个数据集（标识为 *file_name*）。

```
sort RECORD_NUMBER
merge 1:1 RECORD_NUMBER using file_name
```

3.D.i Positions and Statuses 职位和社会地位

除非另有说明，否则下列大部分变量都由软件生成，该软件通过录入户口册中社会地位原始记录的拼音给变量编码。在辽宁八旗人口册中，政府把一个成年男性的社会地位记录（如果有的话）置于他的名字之前（图 7）。这些身份记录通常代表个人与国家有关的信息，一般包括薪酬、官方头衔、荣誉和考试头衔、是否残疾等。有时相关的种族信息也被记录下来，以便对特殊的种族群体进行登记。用户可以使用这些变量来分析个人社会经济地位和政治地位差异的原因和影响。通过附加操作，也可以利用这些变量来构造新变量，来考察个人家庭成员或亲属网络的特点。

BAI_ZONG 百总

BAI_ZONG 是一个标识变量，表示成年男性是否有百总这一职位[33]。 百总是清朝绿色标准军队（green standard army）中的小官员，绿色标准军队的主要军事系统独立于八旗系统。它的等级排名是 7a[34]。一个百总可以享受每月 3 盎司银两[35]的待遇。在 CMGPD_LN 数据集中，0.29%的 20 岁至 50 岁之间的男性观察者有百总头衔。这个头衔的存在表明，虽然 CMGPD_LN 人口在八旗组织控制之下，但是仍然有一些成年男性在绿色标准军队中服役。

BANNER 旗

BANNER 表示个人附属于八旗之下。大多数附属于八旗组织的个人被归为八旗中一旗之下，以旗帜的颜色和设计来区分：正黄旗，镶黄旗，正白旗，镶白旗，正蓝旗，镶蓝旗，正红旗，镶红旗。八旗分为两个等级：第一等级包括正黄旗，镶黄旗和正白旗，第二等级包括其他地位稍低的五旗。

BANNER 与其它身份和地位变量不同，它的出处不在个人登记条目之中，而是与其家庭所在的村庄名字一起记录。因此，在一本户口册中，住在同一村庄并且附属于同一旗的所有家庭成员 BANNER 值相同。没有特定村庄的住户，或者有特定村庄信息却没有八旗注释的住户，其 BANNER 值将被视为缺省。同一个村庄不同的家庭可能会有不同的旗帜背景。

[33] 在介绍清代机构一书中，百总官衔标为 bazong。

[34] 在清朝的军事系统中，排名最高的是 1 b，排名最低的是 7a。

[35] 有关详情请参阅 Luo, Ergang, *Lvying bingzhi,* p. 342。

在此种情况下，不同户口册中的所有家庭被分为不同的组，每个组都标注有自己的旗帜和村庄。

BAOYANG 抱养

BAOYANG 是一个标识变量，表示在户口册中一个成年男性是否可以被归类为抱养。抱养是一个正式的八旗分类，用于确定个人和其后裔的身份。被抱养的个人最初通过非亲属形式被收养，随后他们从平民转化为旗人家属，最终成为世袭旗人。抱养旗人不是正身旗人，并经常与其他奴人（如投降）被分为一类，待遇相当。在 CMGPD_LN 数据集中，只有两个数据库-大辛屯人丁户口册（数据库 12）和开档投从抱养人丁户口册（数据库 28）存在抱养的观察值。使用者应该注意，这种地位与其它大多数地位不同，因为它是一种遗传性的地位。

BI_TIE_SHI 笔帖式

BI_TIE_SHI 是一个标识变量， 表示一个成年男子是否有笔帖式一职。笔帖式是八旗系统中的中等职位，主要处理政府文书工作。根据排名位置，笔帖式享有以现金，大米和豆类形式支付的工资。

DING_DAI 顶带

DING_DAI 是一个标识变量，表示一个成年男性是否被记录有此身份。顶带是一个荣誉头衔，符合此头衔的人可以在他的官帽上被授予佩戴一种装饰品作为“顶带”。顶带可以通过在朝廷任职或购买获得，没有实际官方职位的人也可被授予顶带作为奖励。因此，在我们的人口数据集中，顶带只是表明个人的荣誉头衔。顶带一般伴随着一个官僚等级（RANK），这个等级介于3至9之间，表示官僚地位。在分析中，使用者可以结合PIN变量来衡量一个人的荣誉头衔。

DISABLED 残疾的

DISABLED 是一个标识变量，表示一个成年男性在 POSITION_CODE（或有时候 POSITION_2_CODE）中记录的行政地位是否代表残疾。具体而言，这些行政地位中一般带有“残”或“废”。一个最常见的例子就是“废丁”。

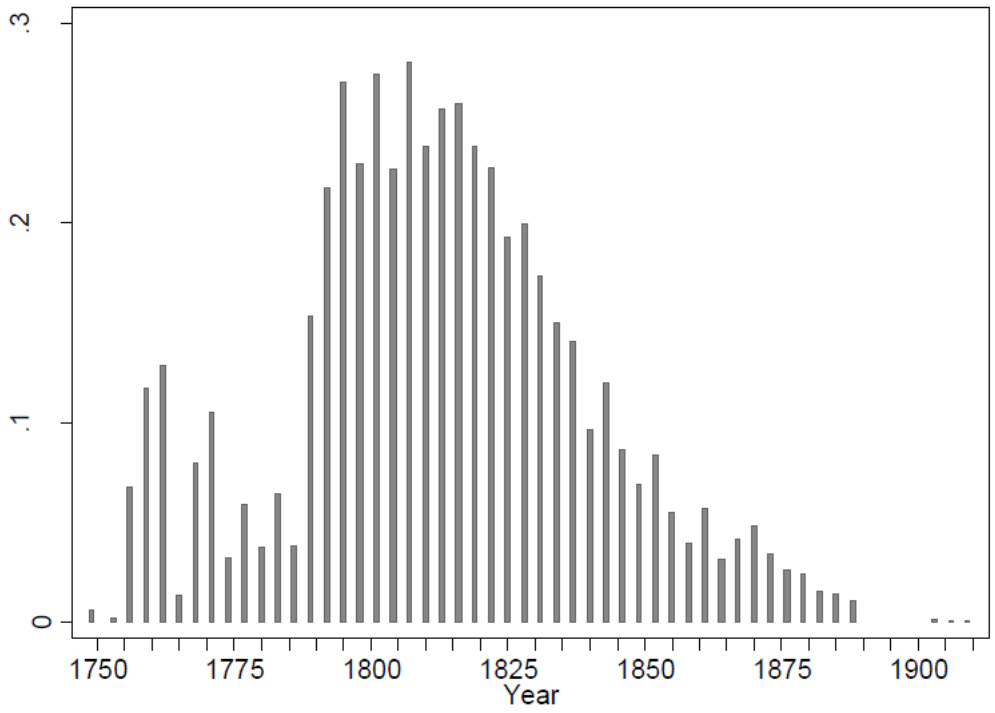
因为这个变量是基于 POSITION_CODE 和 POSITION_2_CODE 中所记录的地位是否含有表示残疾的字符而构建的，当人们的行政地位中没有表示残疾的字符时，这个变量的值不会被设置为 1，即使他们的 DISABILITY_CODE 中有具体的残疾类型。具体残疾类型是 DISABILITY_CODE 的基础，它们在户口册个人条目中被记录的位置与行政地位不同，而 POSITION_CODE 和 POSITION_2_CODE 是根据行政地位构建的。

使用者应该注意，残疾记录随着不同时间、不同数据人群的变化，反映了在记录这个变量时不同的时代和国家利益的制度背景。表 15 的残疾列表显示，在辽宁省成年男性中正身旗人的残疾记录比例比其它旗人要高很多。在辽宁正身旗人成年男性中，约 8.63％的人有残疾记录，但特殊服务人口的成年男性残疾记录比例仅有 4.13％，成年供养人口的残疾记录比例只有 3.8%。这些特殊服务及供养人口的残疾比例较低，可能反映了它们的残疾状态关乎较小的国家利益。

表 15 不同人口分类 20-50 岁成年男性残疾数量

人口分类		残疾的		
		0	1	总计
辽宁正身旗人	频数	312,152	29,489	341,641
	百分比	91.37	8.63	100
辽宁特殊义务旗人	频数	44,989	1,937	46,926
	百分比	95.87	4.13	100
辽宁底层阶级	频数	15,425	610	16,035
	百分比	96.2	3.8	100
总计	频数	372,566	32,036	404,602
	百分比			
		92.08	7.92	100

随着时间的推移，残疾记录也发生了巨大变化。我们强烈建议使用者以年份为单位来检验 DISABILITY 和 DISABILITY_CODE 的分布模式，并在人口分析中选择适当年份的观测值。图 11 总结了编码为残疾的成年男性的比例。



注：Year：年份

图 11 21 岁以上编码为残疾的男性比例，以年为单位

ESTIMATED_INCOME 估计收入

ESTIMATED_INCOME 反映了拥有国家官职的个人的工资收入（以两为单位）。这些值

是从《国朝建业初基》中估算出来的，这本书介绍了 1875 年到 1900 年之间盛京内务府机构的体制，和户口册内记录的的职位或头衔。由于相同职位的薪金在一个时间点上不尽相同，ESTIMATED_INCOME 应被视为一个近似值。如果一个男性在他的社会地位中被记录有多个职位，则将他的工资相加产生一个估计收入。

EXAMINATION 考试

EXAMINATION 是一个标识变量，表示男性是否有学生头衔或者是否在科举考试中取得功名，包括文考和武考。文考头衔包括官学生、生员、举人、进士。武考头衔包括武生和武举。

头衔是通过购买或通过考试而取得的。这个变量可以用来衡量个人和他的家庭在教育成就或教育方面的投资。在 CMGPD_LN 数据集中，约 0.28%的男性或 0.39％年龄介乎 20 至 50 岁的成年男性都有这样一个头衔。

EXPELLED 开除

EXPELLED 是一个标识变量，表示男性个人的记录状态中是否有“开除”标注，如果有就意味着政府已把他从旗人机构中开除，并把他从户口册中移除，因为他犯有某些罪行或违反了某些规则。被开除后，个人并不一定离开居住地。在 CMGPD_LN 数据集中，约 0.07％的男性，或 0.05％年龄介于 20 至 50 之间的成年男性有被开除的记录。有趣的是，约 0.22％的 50 岁以上男性有被开除的记录。

GAO_LI 高丽

GAO_LI是一个表示男性是否有高丽或韩国人记录的标识变量。在CMGPD_LN数据集中，约0.05％的男性被记录为韩国人。然而，CMGPD_LN 29个数据库中只有两个——盖州满汉人丁（数据库8）和棉花盐靛庄头壮丁（数据库25）——有韩国人口记录，占这两个群体中男性人口的0.41％。

DEAD 死亡

DEAD 是一个标识变量，表示男性的状态中是否标注了“故”。通常情况下，有这个注释的个人通常是由国家受薪的，并且已经死了。然而，这个变量并不是显示发生死亡的重要记录，因此不应该被用于分析死亡率。

GUAN_XUE_SHENG 官学生

GUAN_XUE_SHENG是一个标识变量，表示一个成年男性是否有官学生头衔。官学生就读于公立学校，为国家主办的考试而学习。为了在政府学校就读，男性要通过资格考试。这个变量可以用来测量个人的教育程度。用户要注意的是，EXAM 变量也包括具有官学生头衔的人。在CMGPD_LN数据集中，约0.11％的男性具有官学生头衔。

HONORIFIC 荣誉头衔

HONORIFIC 是一个标识变量，表示一个成年男性的荣誉头衔记录。表 16 列举了个人在 CMGPD_LN 中的荣誉头衔。

表 16 作为官方职位修饰的荣誉头衔

敕封	根据皇家的敕命任命

敕赠	根据皇家的敕命赠予
大夫	官僚体系中的一个荣誉头衔
顶带	官方礼帽装饰物
恩禀	
恩赐	受皇家的亲睐赠予
恩贡	受皇家亲睐的官学生
封典	根据皇家的敕命授予
诰封	根据皇家的命令封赏
诰赠	根据皇家的命令赠予
功牌	一块刻着军事荣誉的木片
花翎	官方礼帽装饰物，通常用孔雀尾巴制成，适用于官职在五品或五品以上的官员
郎	官僚系统中一些荣誉等级的后缀
蓝翎	官方礼帽装饰物，通常用雉鸡尾巴制成，适用于官职在六品或六品以上的官员
骑尉	官僚系统中一些荣誉等级的后缀
衔	等级，当用于实际官职头衔时，通常指的是一个荣誉头衔
苑丞	官僚系统中的某些荣誉等级

JUAN 捐

JUAN 是一个标识变量，表示男性的官职是否为购买。从十七世纪后期开始，清政府便出售各种头衔以敛财和补贴军费开支。这种做法一直持续到清末，在十九世纪中叶后尤为盛行。销售的头衔包括实际官衔、荣誉官衔和考试等级。这个变量可以被看做个人以及他的家庭社会经济地位的一个指标，因为购买一个头衔往往需要大量资金。

MAN_ZHOU 满洲

MAN_ZHOU 是一个标识变量，表示男性是否为满族。盛京内务府的旗人大多数是汉族，被称为汉军，或汉元帅。因此，在户口册中，家庭层面上并没有系统的种族记录。通常情况下，种族的信息出现在户口册的标题页上。只有种族不是汉的家庭，种族信息才会标注在户主的名字后。在 CMGPD_LN 人口中，只有 0.06%的男性记录为满族。

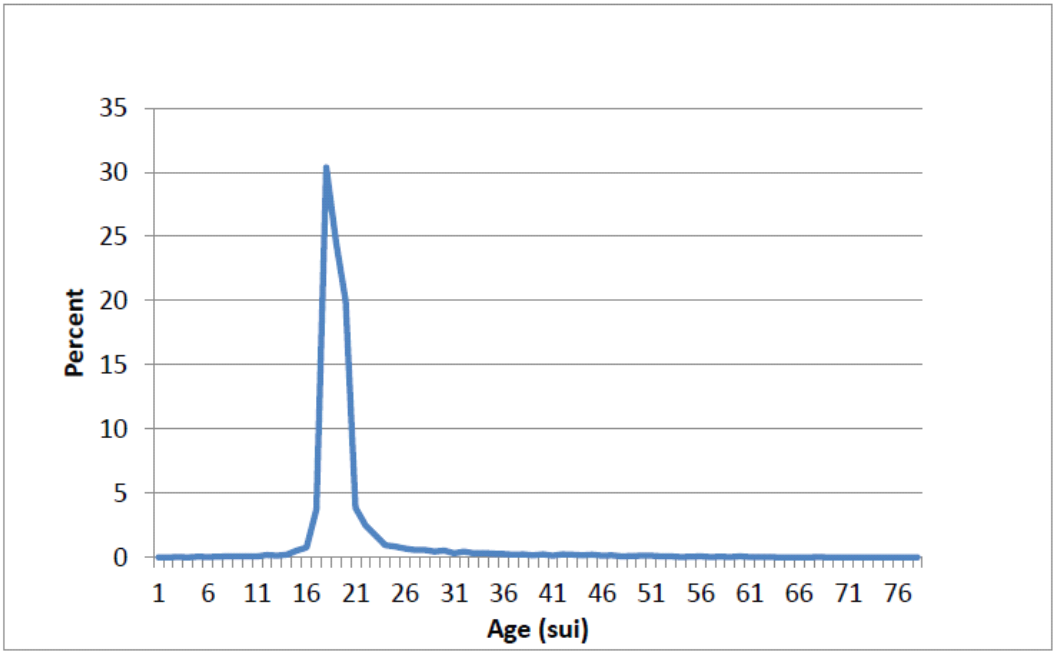
LING_CUI 领催

LING_CUI 是一个标识变量，表示男性是否具有领催的头衔，或税务言教头衔的记录。领催在八旗系统官职中等级较低，负责征税和征收劳务。一个领催一年可以拿到 36 两白银，每月 3 两。

NEW_DING 新丁

NEW_DING 是一个标识变量，表示一个男性是否被新分配为丁，丁就意味着男性开始向国家缴纳税收和提供劳务。在原始数据中，把字符“新”放在字符“丁”前面表示男性的社会地位为新丁。如图 12 所示，虽然很多被标记成新丁的男性年龄范围广泛，但是初始登记为丁的男性主要在 18 至 20 岁之间。74.86%的男性开始被登记为丁，并开始缴纳税收和提供劳务。

图 12 CMGPD_LN 中新丁的年龄分布



注：Percent：百分比

RANK 等级

RANK 是表示男性头衔记录中官僚排名的变量。这个数值从户口册的原始记录中录入。在清代官僚中，排名范围从 1 最高至 9 最低。CMGPD_LN 中的官僚排名范围从 3 到 9。此外，这些排名不仅适用于官衔，也可以用来表示荣誉等级。

PURCHASED_TITLE 购买的头衔

PURCHASED_TITLE 是一个标识变量，表示男性的头衔是否有购买记录。除了含有字符“捐”（表示购买）外，购买的头衔还包括那些含有“监生”（购买政府学生身份）的词。在 CMGPD_LN 数据中，约 0.14%的 18 至 60 岁男性拥有购买头衔。

QIAN_ZONG 千总

QIAN_ZONG 是清朝绿军中较低的官衔。它位于6b[36]。一个千总可以享有每月4两白银[37]的待遇。

SERVICE_DING 服丁

SERVICE_DING 是一个标识变量，表示在户口册中，一个成年男性是否被分配到为内务府提供专业化服务的职位。这些专业化的服务包括收集蜂蜜、养蜜蜂、捕鱼、采摘棉花、制革和印染等。这些服务提供者的社会地位低于常规务农的成年男性，因为他们服务义务较重。在 18 世纪时，大多数这样的成年男性也有以现金支付给内务府代替劳役的选择。

如表 17 所示，虽然服务人群大多集中在为特殊服务人口编制的户口册中，但他们也分

[36] 在清朝军事系统中，最高排名 Ib，最低排名是 7a。

[37] 有关详情请参阅 Luo，Ergang, *Lvying bingzhi* P342。

散在其它大部分都是正身旗人的户口册中。

表 17 不同数据人群中男性服丁分布

DATASET	非服丁		服丁		总计
	观察值	百分比	观察值	百分比	
道义	71,666	99.83	123	0.17	71,789
盖州	25,274	99.97	7	0.03	25,281
打蜜	10,634	55.77	8,433	44.23	19,067
城内	30,983	99.99	2	0.01	30,985
棉丁	17,156	48.61	18,137	51.39	35,293
牛庄	25,440	75.20	8,388	24.80	33,828
肥城	43,140	99.97	14	0.03	43,154
满汉	32,186	100.00	0	0.00	32,186
大甸子	45,061	100.00	0	0.00	45,061
郭三屯	20,691	100.00	0	0.00	20,691
八棵树	30,030	100.00	0	0.00	30,030
大辛屯	53,119	99.96	19	0.04	53,138
年马岱海	19,210	53.04	17,010	46.96	36,220
长寨子	27,847	100.00	0	0.00	27,847
照化屯	31,346	100.00	0	0.00	31,346
貂皮屯	26,997	52.66	24,274	47.34	51,271
郎家堡	28,754	98.94	307	1.06	29,061
王指挥屯	36,262	99.55	164	0.45	36,426
阿尔吉善	9,750	100.00	0	0.00	9,750
海州	72,952	100.00	0	0.00	72,952
王多罗树牲丁	3,392	54.30	2,855	45.70	6,247
王多罗树人丁	13,078	95.19	661	4.81	13,739
瓦子峪	45,454	99.92	36	0.08	45,490
五户	26,558	100.00	0	0.00	26,558
棉盐丁	32,600	70.24	13,809	29.76	46,409
苏拜公	2,655	100.00	0	0.00	2,655
开档	4,669	88.48	608	11.52	5,277
开档投从抱养	17,964	96.26	697	3.74	18,661
棉花盐靛小门	11,696	69.65	5,096	30.35	16,792
总计	816,564	89.03	100,640	10.97	917,204

TOU_CONG 投从

TOU_CONG 是一个标识变量，表示个人或他的祖先是否通过上缴他们所有的财务，包括自己和他们的财产来加入八旗。因此，通过这种方式加入八旗的人地位相当于仆人。此外，这中投从的地位是世袭的。内务府把他们的记录与正身旗人和服务旗人的记录分开，编入另外的奴隶人口（这些人的祖先通过抱养加入八旗）中。在CMGPD_LN数据集中，有投从状态的个人只存在于数据库28中。

XIAO_QI_XIAO 骁骑校

XIAO_QI_XIAO是一个标识变量，表示个人是否有骁骑校一职，在八旗的军事地位中排列中等。一个骁骑校，通常翻译为lieutanent，排名6A，并享有每年60两白银的工资。在当地的旗人社会中，骁骑校的社会地位相当高。骁骑校可以升职为佐领，或一旗将军。

ZHI_SHI_REN 执事人

ZHI_SHI_REN 表示个人是否被记录有该职位。执事人是八旗中的一个附属官衔，级别同士兵相当。一个执事人每年可获 24 两白银。

3.D.ii Family and Household 家庭与户

BIRTH_ORDER 出生顺序

此变量表示个人的出生顺序，首先利用 FATHER_ID 的值找到同父所生的子女，然后对比同父所生子女的出生年得到 BIRTH_ORDER。对于在户口册中找不到父亲或 FATHER_ID 缺失的人，BIRTH_ORDER 被记录为缺省。同一年出生的同胞兄弟姐妹，顺序被随机打乱。根据定义，这个变量要求子女的存活时间足够长以便可以被登记到，不包括那些已经出生，但是在父母还没来得及登记前就死亡的婴儿。因此，BIRTH_ORDER 可能低估了记录中个人的出生顺序。根据 BIRTH_ORDER，先出生的人（如 BIRTH_ORDER= 1），实际上可能是第二个或以后出生的，其哥哥姐姐可能都在登记前死亡了。

原则上，RELATIONSHIP 中的信息也可以提供原始户口册中记录的出生顺序。例如，RELATIONSHIP 为 1s2s 代表一个人长子的第二个儿子。这些信息在此变量的构建中不会被使用。

BIRTH_ORDER_SEX 出生顺序_性别

与 BIRTH_ORDER 相同，除了按照性别分组外。

HH_SIZE 户的规模

这是当前记录中存活个人的计数变量。它基于个人具有相同的 DATASET, YEAR, ZU _SEQ 和 HOUSEHOLD _SEQ 的值，不包括已确定因死亡，婚姻状况，或上次登记后出现其它重要事件而退出户口册的个人。1789 之前户口册观测值的 HH_SIZE 被设置为结构性缺失（-98），因为这些观察值在户口册中并没有区分不同家庭。

HH_DIVIDE_NEXT 下次户口册中家庭分割情况

这是表示当前户口册中一个家庭的个人在下次户口册中也可被观察到的家庭的计数变量。它用于分析家庭分割情况。如果当前户口册中一个家庭的所有个人在下次户口册中还属于这个家庭，HH_DIVIDE_NEXT的值被设置为1。 如果家庭中没有个人能够在下次户口册中被观察到，HH_DIVIDE_NEXT的值被设置为0。1789之前的户口册和每个DATASET中最后一本户口册中的HH_DIVIDE_NEXT被设置为结构性缺失（-98）。

SENIORITY 年长

这是一个数值变量，表示同胞兄弟姐妹之间的年长状况，同胞兄弟姐妹需有相同的父亲（如 FATHER_ID 相同）。户口册中找不到父亲（如 FATHER_ID 缺失）的个人的 SENIORITY 被设置为结构性缺失。一本户口册中最年长的个人，SENIORITY 的值将被设置为 1，其余的兄弟姐妹根据 BIRTHYEAR 计算出来的年份排序。因为它指的是当前户口册中现存兄弟

姐妹的顺序，它可能与 BIRTH_ORDER 差别很大，因为出生顺序会涉及到任何户口册中所有同父子女的观察值。

SENIORITY 可跨家庭进行计算，因此虽然个人在所在家庭的同胞中是最年长的，如在另一家庭中有更年长的同胞，此变量的数值被设置为 2。

SENIORITY_SEX 年长_性别

与 SENIORITY 相同，除了按照性别分组外。

3.D.iii Kin Groups 血亲组

血亲组变量旨在对因为有一个共同的父系祖先而构成联系的个人进行分组观察。这些变量允许不同定义下的父系血亲群体作为分析单位。我们提供一些不同定义下血亲群体的变量，从狭义的定义到广义的定义，最狭义的定义是通过 FOUNDER_ID 定义的，更广义一点的为 FOUNDER_INFERRED_ID，然后是 UNIQUE_ YLHU，最后是 UNIQUE_GROUP。UNIQUE_HH_ID 的构造与以上变量不同。它把不同户口册中生活在相同家庭的个人分为一组，因此，它与本部分所述的其它血亲组变量不可直接相较。

FOUNDER_ID 祖先身份编号

FOUNDER_ID 是根据户口册中个人的最早父系祖先进行设置的。在一个 DATASET 的所有观察值中，具有相同父系祖先的男性会有相同的 FOUNDER_ID 值。与 YEAR 相结合，FOUNDER_ID 可将每年可追溯到相同父系祖先的个人分为一组。妻子和寡妇被设置了与其丈夫相同的 FOUNDER_ID 值。FOUNDER_ID 为每一个人和他们的祖先构建连接，通过把 FATHER_ID 的值从一代串联到更早的一代，直到一个男性在最早的户口册中被定位，这个男性就代表了户口册中最早的男性观察值。相关的程序是很复杂的，涉及许多矛盾的解决或不清楚地方的处理的问题，因此不同的人独立编写程序来实现相同的目标，根据他们如何处理特殊情况，结果会略有不同。FOUNDER_ID 有 25540 个不同的值，对应最早户口册中 25540 支独特父系世系。

FOUNDER_INFERRED_ID 祖先身份假定编号

在一个 DATASET 中，通过处理同一家庭（或者 1789 前相同一户）中男性 RELATIONSHIP 的值，确定共同的父系祖先。每一个不同的 FOUNDER_INFERRED_ID 都可以把能够追溯到拥有共同父系祖先的男人分为一组。原则上，FOUNDER_INFERRED_ID 表示户口册中同一父系祖先的后裔，它应该比 FOUNDER_ID 有更广义的血缘群体定义。

举个例子解释此变量与 FOUNDER_ID 的关系。对于记录在最早户口册中 RELATIONSHIP 为兄弟的男子，他们和他们的子嗣有同一个 FOUNDER_INFERRED_ID 值，因为他们有一个共同的父亲、共同的祖先，这可以从他们之间是兄弟这一事实中推断出来。他们的父亲不必在户口册中被记录，通过男子之间的相互关系就可以推断他们有同一个父系祖先。对于两兄弟，不同个人的后裔将有不同的 FOUNDER_ID 值，因为 FOUNDER_ID 识别的是户口册中具有相同父系血统的人，他们的祖先必须被记录在户口册中。表兄弟、堂兄弟或其它父系亲属的后裔将有相同的 FOUNDER_INFERRED_ID 值。妻子继承她们丈夫的 FOUNDER_INFERRED_ID 值。

因为这个变量在很大程度上依赖于复杂的算法来处理 RELATIONSHIP 中的字符串，从而确定有共同父系祖先的相关个人，因此我们应当把它看作是实验性的，需谨慎使用。实际上有 28,832 个不同的 FOUNDER_INFERRED_ID 值，比 FOUNDER_ID 的值多好几千。然而，由不同 FOUNDER_INFERRED_ID 识别的观察组数量的中位数为 12，比 FOUND_ID 识别的观察组数量的中位数 9 要大。

UNIQUE_GROUP 特殊一组

UNIQUE_GROUP 代表了血亲组最广义的定义。这种计算的目的是找出列于最早户口册中具有相同姓的一户成员，将其假定为拥有共同父系祖先的后裔。这一假设很大程度上是基于相邻一户 的同姓男性往往有相同的特点。UNIQUE_GROUP 有 1063 个不同的值，因此根据构建这个变量的定义，大约 25 万 CMGPD-LN 中的个人可被分为 1063 个不同的亲属群组，同一组的个人是同一父系祖先的子嗣，他们的祖先可能生活于最早一册户口册之前。那些具有不同 UNIQUE_GROUP 值的血亲组，事实上却是更早父系祖先的相同后代的情况也是有可能的，但由于是自动连接程序产生 UNIQUE_GROUP 的值，这个共同的血统可能不会被程序识别出来。

UNIQUE_HH_ID 特殊户编号

UNIQUE_HH_ID 用于记录连续户口册中生活在相同家庭的个人组观察值。由于成员的变更和分家等过程，家庭会随时间而变化，这就需要纵向对家庭的定义进行解释。为了创建这个变量和进行赋值，我们选择了在纵向范围内定义家庭，因为个人可能会通过死亡、出生、婚姻，或其他进程退出或进入户口册，但这与分家不一样，分家会导致同一个家庭中的个人在下次户口册中出现在不同的家庭里。当一个家庭被分家后，我们把这个家庭看作“新”的家庭，并给家庭成员赋予他们自己的 UNIQUE_HH_ID 值，这个值他们会一直保留，直到他们再次分家。至少在原则上，除了家庭成员的变更和其它观察不到的家庭变故，具有相同 UNIQUE_HH_ID 值的个人，在不同年份应该享有相似的的家庭环境。即使当前户口册的户主与下次户口册的户主不同，UNIQUE_HH_ID 保持不变，除非分家。

家庭的其它纵向定义当然也是可能的，并有可能通过编程为不同定义下的新家庭创造 UNIQUE_HH_ID 的替代变量。例如，UNIQUE_HH_ID 的替代变量可以随着不同户口册中同一家庭户主的不同，产生新的值。这个变量的假设是，如果一个家庭在相邻户口册中家庭成员观察值大体相同，但是在第二本户口册中户主发生了变化，我们则认为这个重大的改变可以使其被作为一个新家庭。

UNIQUE_YI_HU 特殊一户

UNIQUE_ YIHU 代表了一个介于 FOUNDER_INFERRED_ID 和 UNIQUE_GROUP 之间的父系血亲组定义。在最早一册户口册中，若最早的父系祖先被记录在同一个一户里，这些个人的 UNIQUE_YI_HU 会被设置为相同的值。一户应该由相关的个人组成，在实践中，我们对数据的检查表明了几乎在所有情况下，相同一户的男子有相同的姓氏，并且通常有其它暗示亲属联系的信息，最明显的是他们的名字具有代际特点。在 1789 年前的户口册中，并没有区分住宅家庭，相同一户的个人都被记录为与一户的户主有亲属关系。与之前相同，妻子和寡妇沿袭了她们丈夫的 UNIQUE_YI_HU 值。总共有 2688 个不同的

UNIQUE_YI_HU 值。

3.D.iv Events 事件

分析版数据提供了一些直接来自原始数据的附加变量。它们的存在主要是为了数据的完整性。

ADOPTED_IN 领养

ADOPTED_IN 是一个标识变量，表示一个男性观察值中是否包括一个曾被领养过的注释。这个注释表示在相邻两册户口册的间隔期间，个人是否被其父母（从当前 RELATIONSHIP 变量的值推断）收养。领养在中国指的是过继，通常发生在有亲属关系的男性中。只有几百人在户口册中有这样的记录。

很可能存在许多未登记的收养案例，因为被收养男孩的年龄还不够大，他们的养父母还不想给他们登记。我们这种猜测的依据是，从已登记男孩的分布与一个简单的生育模型预测出的男孩的分布的对比中，可以看出，没有儿子的已婚男性比例比模型中预测的低。显然，这也不能完全证明。

因此,使用这个变量时应该十分小心。数据中对于收养的记录很可能是不完整的，也不应该被作为完整的记录加以分析。我们将其包含在内是为了数据的完整性，使数据尽可能全面的包含原始记录，即使我们认为原始记录值得怀疑。

ADOPTED_OUT 给别人收养

ADOPTED_OUT 是 ADOPTED_IN 的补充。表示一个男性观察值中是否包括一个曾被送给别人收养的注释。这个注释表示在相邻两册户口册的间隔期间，个人是否被其父母（从当前 RELATIONSHIP 变量的值推断）送给别人收养过。

使用这个变量时也要十分小心。数据中对于收养的记录很可能是不完整的，也不应该被作为完整的记录加以分析。我们将其包含在内是为了数据的完整性，使数据尽可能全面的包含原始记录，即使我们认为原始记录值得怀疑。

EVENT_1 事件一

EVENT_2 事件二

我们将 EVENT_1 和 EVENT_2 包含进来是为了数据的完整性，也是为了方便高级使用者使用。我们建议主要对婚姻、生育和死亡感兴趣的使用者利用 Event 变量。

EVENT_1 和 EVENT_2 是由编码员输入的，表示特定类型的进出户口册事件的数值变量。它的值代表了具体事件。

大部分的值都是根据户口册的注解而设置，这些注解表明了每本户口册涵盖的三年中特殊事件的发生。表 18 总结了对应 EVENT_1 和 EVENT_2 每个值的注解或其他事件。上次户口册中存在而当前户口册中不存在的情况尤其可能被记录在原始户口册中。例如，在一次记

录中，一个人在过去三年中已经死亡，那么他在户口册中的条目注释为“故”。其它最常见的退出户口册的原因有以下几个，对女儿来说是“出嫁”，对寡妇来说是“孀妇改嫁”，对男性是潜逃。出生（6）、入嫁（7），或新出现（9），都由编码员为第一次出现在户口册中的个人观察者设置，其值反映了编码员认为的他们进入户口册的原因。它们没有在原始户口册中以注释的形式出现。我们设置两个事件变量，因为在某些情况下，编码员会为一个事件输入两个值，一个是通过对比之前的户口册判断个人进入户口册事件，另一个是直接记录在户口册中的退出户口册事件。

表 18 EVENT_1 和 EVENT_2 编码解释

事件编码	英文	拼音 （按照记录）	汉字（按照记录）
1	Death	Gu	故
2	Outmarriage	Chu pin	出娉
3	Remarriage	Shuang fu gai jia	孀妇改嫁
4	Absconded	Chen tao	陈逃
5	Migrated out	Qian chu	迁出
6	Birth		
7	In-Marriage		
8	Adoption in	Jiru	过继（入）
9	New appearance		
10	Lost	Mishi	阵迷
11	Adoption out	Jichu	过继（出）
12	Expelled	Fa Qian	发遣
13	Returned after absconding	Chen tao hui	陈逃回
14	Absconded	Tao,baotao	逃，报逃
15	Changed (out)	Bian geng(chu)	变更（出）
16	Changed (in)	Bian geng (ru)	变更（入）
17	Removed from the registers	Xiaodang	销除旗档
18	Became a monk	chujia	出家，陈出家
19	Sent on assignment	Bo chu dang chai	拨出当差
20	Deleted	Shan chu	删出
21	Adopted in	Guo ji gui zong(ru)	过继归宗（入）
22	Adopted out	Guo ji gui zong (chu)	过继归宗（出）
23	Returned after amnesty	Fa qian she hui	发遣赦回
24	Returned to commoner status	xiao dang gui ru min ji	销档归入民籍
25	Registered	Zhu ce (ru ce)	注册（入册）

3.D.v Naming 命名

命名变量都是标识变量，表示男性名字有无特点。户口册中原始登记名字的汉语拼音在限制版数据中以 NAME 变量出现，大部分变量的创建都是基于这些拼音字符串。这些标识变量的值可能在相同男性个人的不同观察值中有所不同，因为男性记录的名称可能会随时间而改变。男孩们通常用相当简单的名字，没有姓氏，在成年早期，尤其是他们结婚后，就用包括姓氏的更正式的名字。一些人在后来的生活中改了名字。

DIMINUTIVE_NAME 昵称

DIMINUTIVE_NAME 是一个标识变量，表示男性的名称（限制版数据中的 NAME 变量）拼音中是否带有“小”或“子”。如果在名字中出现“小”或“子”的一个，就表明这个名字是昵称，例如，小狗子或小胖子。如图 13 所示，在男性小的时候，更可能被记录这样的名字，成年后，就换成更为正式的名字。男子在成年后仍然保留这样名字的情况也不少见。

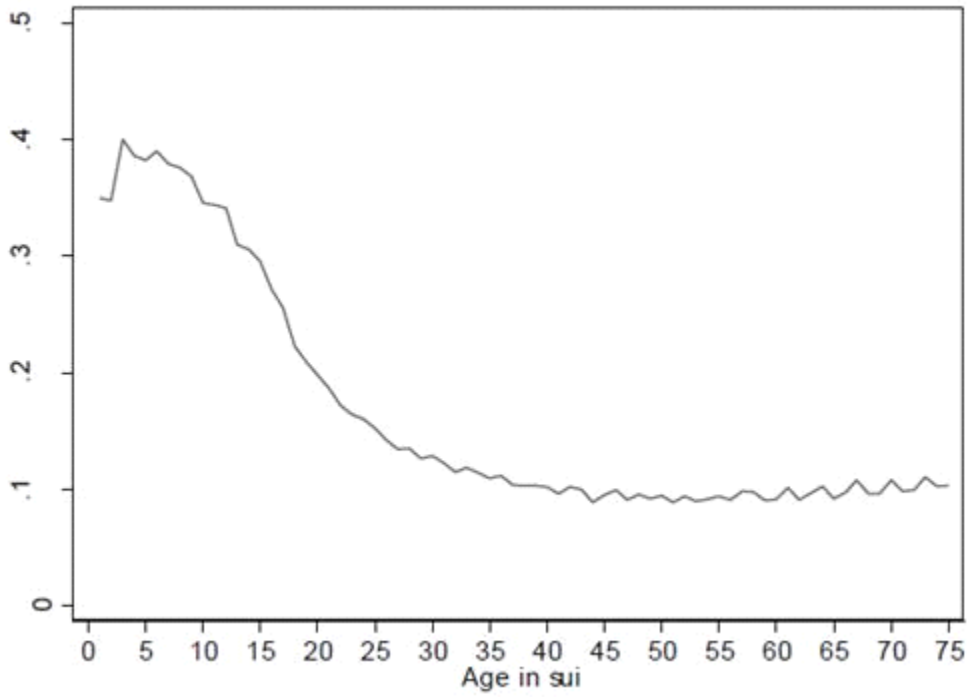


图 13 不同年龄男性有昵称的比例

HAS_SURNAME 有姓氏

这是一个标识变量，表示一个男性是否在户口册中将姓氏作为其名字的一部分。根据图 14，可以看出一个明显的年龄分布。男孩有姓氏记录的可能最小。随着男性年龄的增长，他们更可能有一个姓氏记录。活到 70 多岁的男性百分之八十以上都有一个姓氏记录。从技术层面讲，这是由限制版数据 NAME 变量中名字拼音是否键入空格决定的。编码员把名字从原始的汉字录入成拼音，为了数据分析的需要，对于有姓氏的个人，个人姓氏与名字之间包含一个空格。名字在录入成拼音的时候，字符之间没有空格，因此空格只有在有姓氏的时候才会存在。

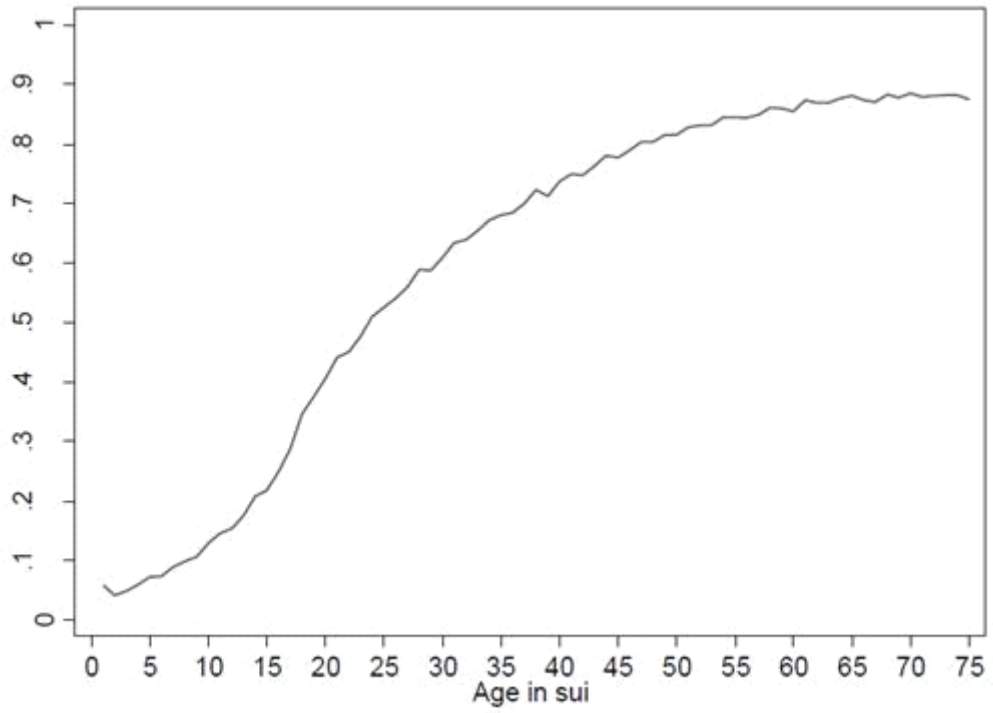


图 14 不同年龄男性记录有姓氏的比例

NON_HAN_NAME 非汉族名字

这个变量表示了一个男性记录的名字是否是非汉族名字。建立这个变量需要几个阶段的工作。首先，我们构建了一个所有出现在数据库中不同男性名字的完整列表。最新版本的清单包括 27746 个不同的名字。我们把名单送到中国东北历史专家的手中，由他们审查并标记公认的非汉族名字。通过多次复查，目前共标识出了 2,626 个非汉族名字。这些名字大多是满族名字。基于结果列表，我们设置了 NON_HAN_NAME 的值。

随着时间的推移，拥有非汉族名字的男性比例下降。图 15 总结了这一趋势。Campbell, Lee 和 Elliott（2002）详细地研究了此趋势，并探讨了家庭背景，社会经济，以及人口因素对男性姓名选择的影响。对这个变量的研究应该考虑到这个趋势。

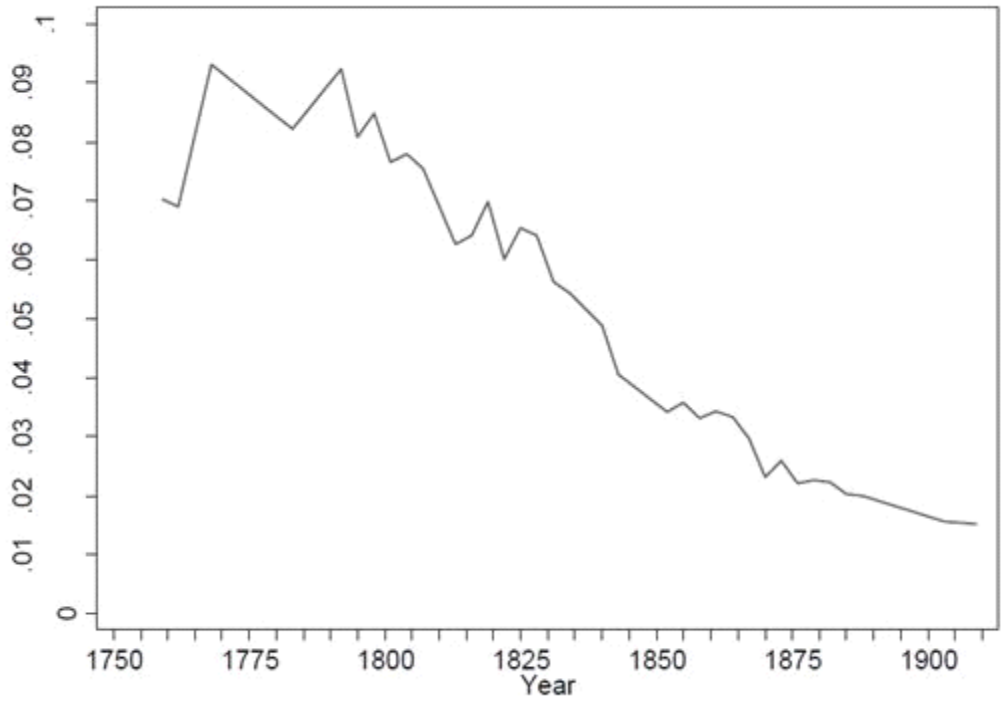


图 15 不同年份有非汉族名字的男性的比例

NUMBER_NAME 数字名字

在这些人口中，男性名字实际是数字的其实并不少见。这是一个数值变量，如果男性有一个数字名字，其值代表了男性名字对应的数字。我们创建这个变量，是为了方便研究父母或个人选择数字名称的值时所遵循的规律。

RUSTIC_NAME 土名

这是一个标识标量，表示男性的名字是否是土名。我们试图创建这个标识变量来识别那些名字暗示了低下社会地位的男性，因为它们的名字包含了动物的名字或者医疗状况。这个定义比 DIMINUTIVE_NAME 的定义要狭窄很多。对 DIMINUTIVE_NAME 而言，如果名称包括“小”或“子”，就被设置为 1。对 RUSTIC_NAME 而言，如果名字包含“狗”、“猪子”、“牛子”、“驴子”、“鬼”、“花子”、“傻子”、“瞎子”、“兔子”、“跳子”、“头子”、“瘫子”这些汉字拼音的话，RUSTIC_NAME 就被设置为 1。我们手动添加了许多不符合这些标准的更多的名字，因为根据我们的调查，这些名子显然是地位低的暗示。根据这些标准，一共有 38426641 个不同的男性名字被列为 RUSTIC_NAME。在我们自己的分析中，经过对这个狭义定义变量的初步调查，我们已经使用了更广泛定义的 RUSTIC_NAME，从而产生了更多的观测值，与社会地位之间的联系也更明确了。

3.E Kinship File 血亲版数据(DS4)

3.E.iAncestry Identifiers 祖先标识

F_ID_1, F_ID_2, F_ID_3, F_ID_4 父亲、祖父、曾祖父、高祖父

F_ID_1，F_ID_2，F_ID_3，和 F_ID_4 分别表示一个男性父亲、祖父、曾祖父、高祖父的 PERSON_ID。对于女儿来说也是如此。对于已婚或丧偶的妇女，它们指的是丈夫的祖先。F_ID_1 与 FATHER_ID 是相同的，F_ID_2 与 GRANDFATHER_ID 是相同的。

这些标识符一举两得。首先，他们使得为个人构建变量描述祖先社会经济成就和特点成为可能。这可以通过一系列的一个或多个 `merge` 操作实现。其次，他们使得把相关个人组织在一起变为可能，以构建描述亲属网络的变量，或采用固定或随机效应来衡量或控制特定类型亲属的共同特征。

F_ID_1 和某些情况下的 F_ID_2，是通过在同一个家庭中加工 RELATIONSHIP 定位一个人的父亲或祖父来构建的。1789 年以前，是通过相同一户中的联系完成的。例如，如果一个人有 `1s1s` 的关系，连接软件就会寻找这个家庭中有 `1s` 关系的人，然后复制他们的 PERSON_ID 作为前者的 F_ID_1。这些操作在每本记录此个人的户口册中重复进行，如果在不同户口册中 F_ID_1 出现冲突，以最早观察到的 F_ID_1 为准。因此，在 CMGPD-LN 中，每人只有一个 F_ID_1，诸如此类。因为有些情况下，用 RELATIONSHIP 识别的父亲可能会在不同的户口册中有所不同，不同解决冲突的方法可能会产生不同值。使用同胞兄弟姐妹信息的比较进行额外处理可以帮助弥补这个缺陷。例如，通过处理 RELATIONSHIP，没有辨识出一个人的父亲，但另一个户口册中，他的同胞兄弟有 F_ID_1，我们便可复制其同胞兄弟的 F_ID_1 的值为此个人 F_ID_1 的值。

F_ID_2，F_ID_3 和 F_ID_4 通过在不同代际重复以上操作而生产。对于每一个人来说，他们父亲的 F_ID_1 被复制成为自己的 F_ID_2。一旦祖父被认定，其祖父的 F_ID_1 被复制为个人的 F_ID_3，等等。在我们自己的分析中，我们又往前推移了多代。我们在此没有把更遥远祖先的标识符包含进来是因为我们想用这四个变量作为数据挖掘潜力的例子。这个复制的过程是很简洁直白的，所以用户应该能够通过 F_ID_4 找到曾祖父父亲的父亲，复制其 F_ID_1 构建 F_ID_5。通过不断重复这个过程可以找到祖先标识符 FOUNDER_ID，也可以根据他们的家谱创造新的组群变量。

只要有可能，我们通过处理 RELATIONSHIP 变量利用堂兄弟，叔伯，或其他亲属信息填补空白。例如，通过处理 RELATIONSHIP 或对他们父亲的 F_ID_1 复制标识，一个人没有定位到祖父，但有一位有表亲在另外一本户口册中有 F_ID_2，该 F_ID_2 的值就被复制了过来。同样，如果通过分析 RELATIONSHIP 定位到家庭里有一位叔叔，叔叔的 F_ID_1 被复制成为个人的的 F_ID_2。

M_ID_1, M_ID_2, M_ID_3, M_ID_4 母亲，祖母，曾祖母，高祖母

M_ID_1 M_ID_2，M_ID_3 和 M_ID_4 分别表示母亲，祖母，曾祖母（父亲的父亲的母亲），高祖母（父亲的父亲的父亲的母亲）。M_ID_1 与 MOTHER_ID 相同。这些变量的产生方式类似于 F_ID_1，F_ID_ 2，F_ID_3 和 F_ID_4。

使用母亲标识符时需注意，根据定义，连接过程是基于户口册中 RELATIONSHIP 所指定的关系，而不是户口册中登记的个人母亲身份的直接信息。如果一个男人在现有户口册中

只有一个妻子观察值，连接系统则会假定这个妻子是他儿子的母亲。如果他儿子真正的母亲从来没有在任何可用的户口册中被记录过，他的儿子会被错误地分配到户口册中妻子观察值名下。由于孀妇再婚是比较少见的，一夫多妻甚至更罕见，我们并不认为这个缺陷对分析有严重影响。

3.E.ii Kin Count Variables (Prefixed by KIN_) 亲属数量变量（前缀是 KIN_）

在血亲版数据里有 KIN_前缀的变量都是当前户口册中不同种类的亲属计数变量。这些计数变量的构建依赖于按照祖先标识符进行分组，然后计算相关个人的数量。例如，兄弟的数量是通过统计当前户口册中具有相同父亲的个人的数量构建的，然后减去 1 用以排除个人自己。同样，表兄妹的数量是通过统计具有相同祖父的人数构建的，然后再减去有相同父亲的人数。第二代表兄妹的数量统计是通过统计具有相同曾祖父的人数构建的,然后再减去有同一祖父的人数。通过进一步细化，我们可以构造具有不同特点亲属的计数变量，包括婚姻状况，相对于指定个人的年长状况，是否与指定个人同住在一个家庭等。

当然，根据 CMGPD-LN 可用的其它任何标准（如职业成就，年龄，生育史等），这些程序可以被推广应用于创造亲属的计数变量上。

与其把所有的变量单独进行命名和讨论，不如解释一下我们遵循的命名变量的标准。表 19 允许我们通过展示一系列变量名称的组成部分，剖析亲属计数变量命名的规则。

表 19 前缀为 KIN_变量的名字组成

性别/婚姻状况	
M	男性
F	未婚女性
MD_F	已婚女性（男性亲属的妻子）
WID_F	丧偶女性（已故男性亲属的遗孀）
年长状况	
O	年长的
Y	年轻的
同居状况	
HH	同居亲属的数量
近亲	
1	同胞兄弟姐妹
2	表兄妹
3	第二代表兄妹
4	第三代表兄妹
2_PA	父亲的同胞兄弟姐妹
3_PA	父亲的表兄妹
4_PA	父亲的第二代表兄妹

3.F Restricted File 限制版数据(DS2)

限制版数据提供个人的姓名和村庄的名字，我们认为这些信息应限于受到研究机构正式许可和承诺，只用数据进行整体量化分析的研究者。我们希望防止个人或商业机构利用CMGPD-LN 进行家谱研究，或者对某一特定名人进行历史研究或传记研究。因为 ICPSR 旨在支持学术研究，我们正在开展学术研究，我们不支持从事上述研究的申请人。若想进行此类研究，可以参考 CMGPD-LN 的原始户口册。本文多次提到，原始户口册在辽宁省档案馆被保存，在 FamilySearch（http://www.familysearch.org）可以找到户口册的影印版。

RECORD_NUMBER 记录编号

这是一个用于合并同一个人在不同户口册、不同年份中的所有观察值的字符串。用此变量合并的细节请参考基本版数据中 RECORD_NUMBER 条目的变量描述。

DATASET 数据人群

见基本版数据中 DATASET 的变量描述。

RAW_VILLAGE_ID 原始村庄号码

ADDRESS 是一个数值变量，它可以与 DATASET 相结合来识别一个条目，这个条目列有编码员录入的汉字和拼音形式的原始村庄名称。不同人口村庄名称的编码都是独立进行的。每次编码员打开 DATASET 中新人口的户口册时，他们会给第一个遇到的村庄名字赋值为 1，下一个新的村名赋值为 2，等等。如果他们遇到了一个已被赋过值的村庄名字，他们就会输入这个值。如果他们遇到了还没有赋过值的村庄名字，他们就会给这个村庄设置一个新的数字。因此，在每个 DATASET 中，不同人群中记录的相同村庄可能会有不同的 ADDRESS 值。因为编码会根据原始户口册中的村庄名字进行设置，即使一个村庄的名字稍微不同于已编码村庄的名字时，编码员也会设置新号码。同一村庄的不同名字被总结在 ASSIGNED_NAME 变量中，ASSIGNED_NAME 和与之有关的 UNIQUE_ VILLAGE_ID 的值跨不同 DATASET 都是一致的。这种整合使记录在不同人口中相同村庄的个人观察值被分为一组。

NAME 姓名

编码员将原始户口册中的个人汉字姓名录入成汉语拼音。

对于男性，户口册要么记录姓和名，要么只记录名。如果姓和名都被记录了，姓氏在先，名字在后。如果姓和名同时被记录，编码员在录入时，按照指示会用空格分隔姓和名的拼音。NAME 中是否有空格是判断原始户口册中是否记载了姓氏的依据，标识变量 HAS_SURNAME表示NAME中是否有空格。

对于女性，户口册中记录的录入成 NAME 的名字取决于她的婚姻状况。住在娘家的女儿记录名字，几乎没有记录姓氏。绝大多数这些名字只是基于出生顺序的简单称谓，例如，三妞或二姐。 其它土名如胖姐或小麻子。在极少数情况下，女儿会被记录稍复杂的名字，如白云或凤英。我们在女儿命名方面还没有作出任何系统研究。

对于已婚和丧偶妇女，户口册记录了她们的娘家姓，以“氏”结尾。加上丈夫姓氏信息，这使得基于名字的选型交配研究有了依据。例如，我们可以研究同姓男女之间结婚的普遍程度。原则上，这在古代中国是避讳的，但我们自己的研究显示，这种现象也偶尔发生。

女性的娘家姓结合其他信息（如女子首次记录结婚的年份和她 BAZI_YEAR、BAZI_MONTH、BAZI_DAY、 BAZI_HOUR 给出的生日），理论上，我们可以将外嫁到其它家庭中的女儿同其娘家相连，但是我们操作的结果不太令人满意。一个问题是，许多女儿没有在户口册中列出，从而无法与有夫之妇相连。只有在十八世纪末和十九世纪初，女儿的记录量十分可观。另一个问题是，我们输入的户口册没有覆盖所有为儿子找到新娘的家庭人口。它们有可能是我们还没有输入的户口册中相同或附近社区中的家庭。也有可能因为与八旗不相关而没有在任何户口册中被记录过。

最初我们把名字录入成汉语拼音，是因为我们开始输入数据时使用的电脑没有便捷的中文输入和储存能力。后来，当我们可以便捷输入中文时，我们决定继续使用汉语拼音而不是汉字。我们主要有两方面的考虑。一是考虑到汉字编码系统的多样性，我们可能会遇到无论选择哪种格式录入汉字，也许都不能和我们或他人将来使用的任何数据处理系统兼容的情况。一个更大的担心是，因为不同户口册中不同的同音字，一个人的名字可能会在连续户口册中书写上不一致。尤其是非汉族名字，他们的名字是满族，蒙古族或其它名称的音译。根据名字搜索个人的有效方法是利用拼音来确保一个人名字的所有同音字也被包括在这次搜索内。因此，我们决定保留用拼音输入的条目。

SURNAME_YIHU 一户姓氏

这个变量根据个人一户或家庭的姓氏推断个人的姓氏。一户应该包括拥有共同父系祖先的男性家庭成员所在的相关家庭。因此，相同一户的男性都应该有相同的姓氏。 从 HAS_SURNAME 的表中明显看出，并非每一个男性都记录了姓氏。该变量是通过一个程序构建的，该程序可检查每个一户中已有的男性姓氏记录，识别出姓氏记录的模式，把它设置为一户 中每个人的 SURNAME_YIHU 值。在绝大多数情况下，一户姓的模式与一户中记录的所有男性姓氏是统一的，但在设置姓的模式时也发现了少数不统一的情况。

ASSIGNED_NAME 赋予的名字

我们手动查看了原始户口册中记录的所有村庄的名字，将其整合成 ASSIGNED_NAME。当户口册记录的名字不同但经我们手动检查确认是相同的村庄时，我们从中选定一个村名来作为该村庄所有个人观察值的 ASSIGNED_NAME，即使记录的村名在不同户口册上略有不同。在大多数情况下，不同的户口册中村名的差异很小，并且从随机检验中可以看出，所有不同版本的名称很明显指向同一个村庄。在许多情况下，差异反映了一个村庄名字中的某个字可以用同音字替代。在一些情况下，名称的差异略大，经过当地地方志和其它档案资料的检验，识别出这些村庄改过名字。

LATITUDE 纬度

LONGITUDE 经度

通过大量工作和劳动，包括仔细考证当代地图和历史地图，向历史学家和当地官员请教，翻阅当地地方志、档案资料，和研究当地历史的学术著作，我们试图手动地定位户口册中尽可能多的村庄。图 1 展示了我们已经手动定位的所有村庄。虽然我们的很多定位都相当精确，尤其是对一直延续至今的较大的村庄，一些今天已经消失的或者被吸收到附近乡镇城市的规模较小的村庄，我们只能近似的定位。对于一些已经消失的村庄，或在标识经纬度的当代地图上找不到的，我们主要根据历史文献上记载的它们与我们已定位村庄的地理关系来近似定位。经度和纬度缺失或表示原始户口册中没有村庄记载，或表示原始户口册有记载，但我们无法定位。

SUOSHU 所属

SUOSHU是一个标识变量，表示当个人是“所属各屯”的成员时，其所在的一户记录的村庄名。在这种情况下，记录的地址和相关的村名指的不是一个特定的村庄，而是一个与叫这个名字的村庄有着行政联系的更大的区域。因此，标记为SHOSHU的观察值不应被包括进任何试图利用实际村庄地址的分析中。“所属各屯”一般出现在最早的户口册中，在这些户口册中，具体村庄姓名的记录比较少。

4 CMGPD-LN Supplemental Files CMGPD-LN 补充文件

4.A Disability Supplement 残疾变量补充说明(DS5)

DISABILITY_CODE 不同值代表的残疾类型的汉语拼音见残疾变量补充说明中的CONDITION_PINYIN。每个 DATASET 中的残疾编码是单独设置的，因此不同 DATASET 中相同 DISABILITY_CODE 的值对应不同的残疾类型。为了使其它版数据的观察值同特定残疾状况的汉语拼音联系起来，我们需要同时根据 DATASET 和 DISABILITY_CODE 合并数据，而不是只根据 DISABILITY_CODE。

在我们自己的研究中，我们为 lao zheng 等具体情况的存在创建标识变量，以及为不同状况分组创建分类变量，都是以汉语拼音为基础的。总的来说，比较简易的做法是在残疾变量补充说明中创建这样的标识变量，然后把它们合并到主文件中，无需把原始的残疾变量也合并进去。

由编码员录入的表示残疾的原始汉字在分析版数据附录中可以找到。

4.B Position Supplement 职位变量补充说明(DS6)

不同POSITION_CODE 值代表的官职和社会地位的汉语拼音见职位变量补充说明中的POSITION_PINYIN。为了使这些变量同基本版或其它版数据的个人观察值联系起来，我们需要将DATASET和POSITION_CODE合并起来。每一个DATASET中，POSITION_CODE的值是单独设置的。

在我们自己的研究中，我们为具体情况的存在创建标识变量，以及为不同状况分组创建分类变量，都是以汉语拼音为基础的。分析版数据中的许多变量都是通过搜索汉语拼音中特殊的字符串而创建的。总的来说，比较简易的做法是在职位变量补充说明中创建这样的标识变量，然后把它们合并到主文件中，无需把原始的职位变量也合并进去。

由编码员录入的职位的原始汉字在分析版数据附录中可以找到。

4.C Position 2 Supplement 职位变量补充说明 2

职位变量补充说明中的 POSITION_2_PINYIN 为其不同值对应的官方职位和地位提供了

汉语拼音。为了将这些官方职位和地位同基本版或其他版本的数据中的个人观察值联系起来，可以将 DATASET 和 POSITION_2_CODE 进行合并。POSITION_2_CODE 的值在每个 DATASET 中是独立设置的。相同 POSITION_2_CODE 的值在不同 DATASET 中可能指的是不同的地位，相同的官职或地位在不同的 DATASET 中可能有不同 POSITION_2_CODE 值。DATASET 与 POSITION_2_CODE 相结合可以指定一个具体的官职或地位。

在我们自己的研究中，我们为具体情况的存在创建标识变量，以及为不同状况分组创建分类变量，都是以汉语拼音为基础的。分析版数据中的许多变量都是通过搜索汉语拼音中特殊的字符串而创建的。总的来说，比较简易的做法是在职位变量补充说明中创建这样的标识变量，然后把它们合并到主文件中，无需把原始的职位变量也合并进去。

由编码员录入的职位的原始汉字在分析版数据附录中可以找到。

4.D Monthly Grain Prices 每月粮食价格

在 18 和 19 世纪，中国有一个全国性的粮价报告系统。每个月，官员们总共报告 12 个价格：他们管辖区内六种粮食的最低市场价格和最高市场价格。这六种粮食是：高粱, 大豆, 小麦，大米，黍米和小米。汇报时使用的单位是每石多少两银子。两是重量的一种计量单位，一两大约等于 38 克。石是粮食体积的测量单位。数据中每月粮食的价格来源于奉天府官员呈交到北京的记录文件，其中包含了户口册涵盖的辽东地区。原始价格是按照农历月份被报告的。James Lee 在 1985、1986 和 1987 年访问北京第一历史档案馆期间收集了 1200 份记录文件 (Lee and Cmapbell 1997, 28)。Wang Yeh-chien 根据他自己在台北故宫博物院中的研究提供了 300 份记录文件。

我们在自己的研究中充分利用了这些价格。Lee 和 Campbell (1997, 27-33)描述了我们自己对每月粮食价格波动研究的结果，以及粮价与气候变化的关系。 Lee, Tan 和 Campbell (1992) 特别集中于研究粮食价格波动和人口行为的关系，并对价格序列做了进一步的讨论。Lee 和 Campbell (2005)查看了价格波动与人口行为关系的长期走势。Bengtsson, Campbell, 和 Lee (2004) 以及 Tsuya, Wang, Alter 和 Lee (2010)对于中国价格的研究成果都是基于这些数据。

皇家粮食价格汇报系统已经是大量文献的研究课题了。例如， Rawski and Li eds. (1992) 的许多研究就是利用同一个汇报系统的粮食价格序列来研究中国的其它区域。Li (2000)通过对不同地区粮食价格序列的计量经济学和时间序列研究，查看了中国北部市场的整合和解体。

在这个文件中，记录有 12 个价格中的每一个与年份的组合。换句话说，12 个记录代表了每年的粮价。为了我们自己的研究和数据的发布，我们将农历月份的价格转换成了阳历月份的价格。

由于户口册是三年一册的，必须将这些月份数据加以处理，它们才可以兼容进行分析。我们希望这些月份数据可以对研究时间序列、计量经济学或可以利用此数据的其他领域的专家学者有帮助。

YEAR 年份

这个变量代表价格的阳历年份。

GRAIN_TYPE 粮食类型

GRAIN_TYPE 代表记录中描述的粮食类型。这里有六种粮食类型：高粱, 大豆,小麦，大米，黍米和小米。

HIGH_LOW 高低

HIGH_LOW 指的是价格记录是地方高价或地方低价。在我们自己的研究中，我们大致假设高价记录往往是消费者价格，是城市或其它将粮食作为消费品的人口密集区的市场价。我们假设低价序列是生产者价格，是农民出售他们收成的乡村的市场价。因此，我们大致假设低价序列往往反映的是乡村的情况，因此与户口册中涵盖的农村人口最为相关。在我们自己的研究中，例如 Lee 和 Campbell (1997),我们发现高粱的低价往往与人口行为大有关系。

MONTH01-MONTH12 月份 01-月份 12

MONTH1 到 MONTH12 变量代表粮食系列（由 GRAIN_TYPE 和 HIGH_LOW 指定）的价格（每石多少两）。尾数 1 到 12 对应了一月到十二月。使用者若想将数据转化成月份系列，可以为每个月创建一个观察值。阳历月份价格根据农历月份价格取加权平均数，权重根据与每个农历月份重合的具体的天数而定。例如，如果阳历四月被农历三月和四月平分，阳历月份的价格应该是这两个农历月份价格的平均数。我们依据一张列表来计算权重，这张表列有数据涵盖范围内每个农历月份第一天的阳历日期。

4.E Annual Grain Prices 每年粮食价格

这个文件中的各个观察值代表了特定年份的年平均价格。我们通过每月价格构建年平均价格。对于每个 GRAIN_TYPE 和 HIGH_LOW 的组合，年平均是 12 个月价格简单的算数平均数。反过来说，这些年平均数是三年平均数的基础，我们利用三年平均数研究价格对人口行为的影响。我们也利用年增长率，在每年的基础上做其与出生率的相关性分析。对于这些年平均数的计算，我们创建了一个版本的每月平均数，在这个版本中，缺失值通过仿样函数被插入。缺失月份的分布和仿样函数的趋势，在与每月和每年价格序列的比较中很显著。我们希望时间序列分析或计量经济学领域的相关专家可以在他们自己的研究中利用月份时间序列构建他们自己的序列。我们提供这些年平均数是为了数据完整性的考虑，更重要的是促进数据的复制。

YEAR 年份

这个变量代表价格的阳历年份。

LOW_RICE, LOW_HUSKED_MILLET, LOW_UNHUSKED_MILLET, LOW_SORGHUM, LOW_WHEAT, LOW_SOY 低价大米、低价黍米、低价小米、低价高粱、低价小麦、低价大豆

HIGH_RICE, HIGH_HUSKED_MILLET, HIGH_UNHUSKED_MILLET, HIGH_SORGHUM, HIGH_WHEAT, HIGH_SOY 高价大米、高价黍米、高价小米、高价高粱、高价小麦、高价大豆

每个变量代表了特定 GRAIN_TYPE 和 HIGH_LOW 组合的平均价格（每石多少两）。

4.F Yingkou Customs Records 营口海关记录

在 19 世纪中期，营口成为可以进行国际贸易的通商口岸之一。1858 年，《南京条约》最先开放了牛庄进行贸易，但是从 1861 年之后，营口成为通商口岸。通商口岸成为出口黄豆、黄豆粕和黄豆油主要中心。营口海关局于 1864 年成立，它也是中国东北第一个海关局。海关局从 19 世纪末到 20 世纪初每年的记录在辽宁省档案馆都能找到，同时还被辽宁省地方志办公室复制了。我们将可以得到的记录录入到一个文件中，以供想要研究这些时间序列反映出来的人口或其他行为与经济状况关系的研究者利用。

营口与南部和中南部村落相隔较近（见盖州、牛庄和其它数据人群涵盖的 REGION 变量）。

YEAR 年份

这个变量代表海关记录的历法年份。

EXPORT_BEAN 出口豆类

这个变量代表出口的豆类总量（以千担为单位）。担是古代中国的一种重量计量单位。一担正常等于 100 斤，大约是 60 千克或者 132 磅。[38]因此 100 担应当等于 60 吨。

EXPORT_BEAN_CAKE 出口豆粕

黄豆粕常被用作其它植物的肥料。这个变量代表出口黄豆粕的数量（以千担为单位）。

EXPORT_BEAN_OIL 出口豆油

出口黄豆油的总数量（以千担为单位）。

EXPORT_BEAN_ALL 出口豆类产品

出口豆类、豆粕和豆油的数量（以千担为单位）。

IMPORT_VALUE 进口价值

所有进口产品的总价值（以多少两银子为单位）。两是重量的计量单位。一两大约等于

[38] 在清朝, 1 斤大约是 596.82 克。

38 克。

EXPORT_VALUE 出口价值

所有出口产品的总价值（以多少两银子为单位）。

TOTAL_VALUE 总价值

所有进口产品和出口产品的总价值（以多少两银子为单位）。

4.G Natural Disasters 自然灾害

辽宁省地方志办公室出版的地方志根据历史记录编纂了自然灾害的列表。这些灾害包括洪水、干旱和地震。我们整理了各种地方志中的信息，将它们按照 YEAR 整理到一张表中，这样就可以与户口册连接用以分析。我们还没有在自己的研究中利用过这些信息，也没有评估过它们。我们提供这些信息以便使用者做探索性分析。[39]

除了文件中描述的灾害，Lee 和 Campbell (1997, 34)总结了我们目前找到的气候事件的其他信息。我们在这里复制这些数据方便使用者将它们录入到时间序列中。

根据 Wang (1988), 1774、 1776、 1778、 1782、1783、1785、1786、1789、1793、1813、1815、1825、1830、1831、1832、1833、1835、1836、1837、1838、1841、1846、1856、1857、1866 和 1869 年份中曾出现反常寒冷的夏天。寒冷的夏天造成农作物生长期缩短，因此这种气候总体上会出现不好的收成。见 Feng, Li 等(1985)对中国东北寒冷夏天的更多细节。

根据 Zhang Peiyuan、Zhang Jinrong 和 Gong Gaofa 提供给我们的沈阳地区（奉天）气候和收成报告,在 1788、1789、1802、1807 和 1815 年存在一些反常的坏收成，被评价为 *shaolian* 或 *lianbo*。在 1794、1795、1796、1797、1798、1799、1801、1804、1816、1818、1820、1822、 1825、1826、1846、1864 和 1873 年也有一些反常的好收成被评价为“较好”、“丰年”、“丰收”，在一些案例中也被评价为“甚好”。

YEAR 年份

事件发生的年份。

---

[39] 海关记录的来源是《辽宁省志-对外经济贸易志》，辽宁省地方志编纂委员会办公室主编，辽宁人民出版社，2003.10. 书中表格：《1864-1911 年牛庄港出口贸易额统计表》，第 10 页， 《清末牛庄大豆、豆饼、豆油出口数量统计表》，第 12 页和《辽宁省志-海关志》，辽宁省地方志编纂委员会办公室主编，辽宁人民出版社，2002.12 书中表格：《1864-1931 年山海关监管营口港进出口货物统计表》，第 62 页。洪水和干旱的信息来源是《辽宁省志-水利志》，辽宁省地方志编纂委员会办公室主编，辽宁民族出版社，2001.12 自然灾害记录：无表格，为记述，P51-76。地震信息的来源是《辽宁省志-地震志》，辽宁省地方志编纂委员会办公室主编，辽宁科学技术出版社，1996.1 书中表格：《2-1985 年辽宁省地震简目》，P70。

DAY, HOUR and MINUTE 日期、时刻和分钟

DAY, HOUR 和 MINUTE 仅仅对地震适用。我们提供这些变量以便使用者将这些记录与其他来源的地震记录结合起来。我们假设位置、日期、时刻和分钟可以精确的识别一场地震。

EVENT_TYPE 事件类型

地方志提供了三种自然灾害的信息: 干旱、地震和洪水。干旱被编码为 1，地震被编码为 2。地方志将洪水分为特大洪水和重大水灾。我们将前者编码为 3，后者编码为 4。两种洪水之间区分的标准尚不明。

DISTRICTS_PINYIN 地区拼音

地方志列举了受灾的行政区域。我们将这个列表录入成了拼音。

WATERSHEDS_PINYIN 分水岭拼音

地方志列举了受洪灾的分水岭，我们将其录入成了拼音。我们将这些信息包含进数据库以便使用者连接到水文或空间数据，将它作为使用受灾地区列表的另一种选择。

EPICENTER_PINYIN 震中拼音

震中地名的拼音。

EPICENTER_LONGITUDE, EPICENTER_LATITUDE 震中经度、震中纬度

震中的经纬度。

EARTHQUAKE_MAGNITUDE 地震强度

用 MS 震级标准估计地震的强度。使用者应当注意这些估计值来源于历史记录。

DISTRICT_01-DISTRICT_13 地区 01-地区 13

识别每个 DISTRICT 是否在受灾名单上的标识变量。

MONTH_01-MONTH_12 月份 01-月份 12

识别一年中的每个月份是否在事件发生月份名单中的标识变量。

SEASON_1-SEASON_4 季节 01-季节 04

识别每个季节是否在旱灾发生期间的标识变量。1 代表春季，2 代表夏季，3 代表秋季, 4 代表秋季。

5 中国多世代人口数据库-辽宁部分（CMGPD-LN）人口分类

尽管所有的户口册都有一致的形式，但八旗的行政管理机构对不同的所辖人口类别实行分开的登记，来体现其所含群体不同的社会政治地位和权利。依据犹他州家谱协会的编目习惯，数据库中共分成 5 个群体类别：登记为正身旗人的普通旗人人群；登记为有特定义务旗人的普通旗人——他们需要为皇家内务府提供特定的服务或是劳动产品；庄园旗人或是其他被奴役的人群——他们或者效忠于皇家内务府，或者是作为被皇家内务府没收充公的原旗人贵族或高管家中的人力财产；登记为在不同军队及相关机构的人群；以及其他各种情况的人群。不同人群间的差异不仅仅可以直接从个人特定的头衔和人群所属的登记类型来鉴别，还可以从其内部的社会行政组织，个人在科举考试中的平均水平和达到的高度，还有从事非农业型的政府岗位的人群比例来辨别。总体来讲，每一类别人群都是很独特的，尽管也出现随着年代的推移，不同人群间的界限变的相对模糊，甚至像 26 号和 29 号数据库所属人群最后出现重新归类的情况。在使用指南的这一部分里，我们主要介绍前三类人群——正身旗人，特殊义务旗人，庄园主和奴役旗人——他们几乎组成了整个 CMGPD-LN 人群——的主要特点，而在第五部分中再具体讨论。

5.A 正身旗人户口册

CMGPD-LN 中 29 部户口册中有 16 部登记的是正身旗人人口。这些正身旗人的户口册在最初是以“某地人丁户口册”命名，后来又改为更复杂的“某地汉军人丁户口册”“某地满汉人丁户口册”等等，但有一些仍旧保留“某地户口册”的形式。我们从由犹他州家谱学会保存并编目的户口册中共鉴别出一系列的 50 种不同的正身旗人户口册，并对其中编号为 1, 2, 4, 5, 7, 8, 9, 10, 11, 12, 14, 15, 17, 18, 20, 22 ,24 的数据库以及数据库 23 的一部分进行了深入研究[40]。

正身旗人由旗内首领（佐领）领导。每一个成年男性（丁）拥有一定的由国家分配的土地，然后每年向皇家内务府上缴不断修改的税赋（Ding, Guo, Lee, and Campbell 2004, 80-85），还包括一些日用品：制衣的布匹，储存的肉食或者其他一些林产品如蘑菇等（Ding, Guo, Lee, and Campbell 2004, 82）。此外，作为正身旗人，他们可以作为工匠或士兵为军队服务——理论上讲是其义务，但实际上是他们最主要的收入和政治权利的来源。

同特殊义务旗人和庄园旗人相比，正身旗人有更多的机会来扩大个人及其家族的财富和权力。如表 20 中显示的，在皇家内务府管辖下的辽东旗人人口，和其他正身旗人人口类似的，都有机会参加由国家举办的科举考试，若某人表现优异，他便有机会成为国家指派的地方官员。相对比的是，庄园旗人则没有参加国家科举考试的机会[41]。因此，我们从这些正身旗人的户口册中可以看到许多在官府任职的，或者作为替补官员，士兵，工匠，以及他们所拥有的各式荣誉头衔。还有一点需提的是，正身旗人人口同其他类型人口相比，所生活的地域流动性更大，不管是合法迁徙还是不合法的迁徙。

[40] 犹他家谱学会目录中旗人群体完整列表请见附件 A。

[41] 根据 1891 年 Yulu 的回忆录,一份内务府 *zhuangding 是可以参加科举考试的，被庄园管理人员管理的庄园旗人却不能* (*Manzhou jiuguan diaocha baogao*, 45-46)。我们感谢 Yizhuang Ding 给我们提供这条信息。

表 20 盛京内务府户口册中正身旗人、特殊义务旗人和庄园旗人比较

	Regular Banner Population	Special Duty Banner Population	Estate Banner Population
人口状态	常规旗人	常规旗人	常规旗人
组织形式	按照另户和父系族(一户或族) 组织	按照另户和父系族(一户或族) 组织	按照父系族(一户或族） 组织，但无住宅户的信息
带俸禄的当地旗人官职	佐领	佐领	管领/~达
监督者	无	庄头或窝棚头	庄头或园头
行政长官	族长	族长	监管者
军事长官*	是	否	否
工匠**	是	否	否
地方官员***	是	否	否
荣誉头衔^	有	无	无
参加国家举办的考试	是	否	否^^
地理流动	高	低	变动大
国家雇佣机会	高	低	低

* 旗人中的军事长官主要指的是八旗系统下拥有士兵或长官职位的人，包括各种各样的士兵头衔（甲丁、 兵，或甲）、委领催、（牛录催长或食俸催长），领催、馆夫、笔帖式、执事人、骁骑校、库使、牧长等等。虽然这些头衔中许多是军事头衔，这些职位实际上是当地旗人政府的日常职位。

** 在CMGPD-LN中，工匠职位是国家服务人员的一种。工匠享有月薪，其总收入比普通农民要高。CMGPD-LN中的工匠职位包括工匠、稞子匠、染匠、花儿匠、裁缝、皮匠等等。

***在CMGPD-LN中，一些人也拥有当地政府系统的官方头衔。大部分这些头衔属于县郡地方官员或衙门。这些职位包括知府、知县、同职、通判、府经历、主事等等。

^ 荣誉头衔是政府授予用于显示个人卓越贡献的头衔。荣誉头衔既可以奖励给官员，又可以奖励给平民。这些头衔一般由一个数字等级和一些荣誉物品（顶带、翎、军功）组成。

^^ 虽然清政府禁止服务人口参加科举考试，国营农场的一些人依旧设法取得考试的权利(Ren 2003; Ding, Guo, Lee, and Campbell 2004)。

普通旗人口与庄园旗人口在行政和社会结构组成上便有差异。管理某一正身旗人群体的最底层行政长官是佐领，而其社会组成是以另户和族为单位的。如图 1.1 所示，在户口册中，户组称为一户，而其下的某一住户则称为另户。户组（族）是作为佐领和每一住户之间的行政单位；根据户口册，我们发现每一另户都在户组的首领的管理下。而在下面讨论奴役和服务旗人人口时，就会发现他们的户口册中并没有另户的区分，而只有户组（一户）。

5.B 特殊义务旗人户口册

CMGPD-LN 29 个数据库中有 8 个数据库是属于特殊义务旗人群体的，包括第 3,5,6,13,16,21,25/29，还有 1830 年之后的第 26 号。这些特殊义务旗人户口册最初的标题为“某地某特殊义务人丁户口册”，到 1864 年以后改为“某地某特殊义务汉军人丁户口册”，也有部分简化为“某特殊义务户口册”。据我们目前的了解，可以确定这些在称谓上的区别并不意味着他们有制度上的差异。我们从犹他州家谱学会编目的户口册中共鉴别出一系列 28 部特殊义务旗人户口册，其中还包括一部没有标题的[42]。同正身旗人人口类似，这些特殊义务旗人也享受国家给予的免费土地，但他们中的所有成年男性按规定需要为皇家内务府提供一定的服务：如采蜜，养蜂，打渔，拾棉花，制革，丧葬等[43]。自 18 世纪开始，他们还有一种选择，即提供这些劳动服务向皇家内务府来换取钱。

尽管国家将特殊义务旗人仍归类为一般旗人——正身旗人，但他们所拥有的生存机会却和庄园旗人类似。从我们对户口册的观察看，他们中基本上没有作为官员，士兵和手工艺匠的个人，同样也没有人有任何从国家科举考试中获得的头衔。因此，与辽东地区的普通旗人相比，他们的社会地位更低，分布的流动性也更低。此外，和庄园旗人相同的是，国家会指派监管人员来管理这些特殊义务旗人，这些人员被称作庄头或是窝棚头。

但单纯从户口册中看，这些特殊义务旗人内部的行政社会结构同普通旗人是一样的。最底层的人口管理人员称作佐领。户口册也是以住户（另户）和户组（一户，族）为单位进行组织的，其中族作为佐领和住户之间的管理中介。这些特征都显现出，尽管特殊义务旗人的地位要比普通旗人要低，但其仍旧属于正身旗人的类别。

5.C 庄园旗人户口册

在 CMGPD-LN 的三种人群中，庄园旗人和其他被奴役人群是最为特殊且最为贫穷的[44]。在 CMGPD-LN 数据库发布之前，许多学者都猜测辽宁地区的旗人人口是“粮庄”或是“官庄”，字面意思即生产粮食的庄园和国家的庄园，而留存下来的这些户口册中的人群都是来自于这些庄园（Isett 2007;Tong and Guan 1995）。但事实上，尽管在辽东地区有 89 座或更多的庄园，我们了解到这些庄园都是受八旗 Board of Revenue 管理，而非盛京皇家内务府，因此这些留存的户口册中记载的人群不应来自这些庄园[45]。实际上，皇家内务府也控制了许多充公的庄园人口和契约奴役人口，他们作为皇室的财产而并非国家的财产。附录 B 中列出

[42] 这些标题的完整列表见附件 A。

[43] 大部分特殊义务旗人有两个来源。一部分是清朝之前八旗的农民，他们可以使用土地但没有土地所有权；另外一部分向八旗捐献了自己土地的农场主，他们靠捐献土地成为旗人。那些有土地使用权的只需要缴纳人头税而不缴纳土地税，而捐献了土地的缴纳土地税但不缴纳人头税(Ding, Guo, Lee, and Campbell 2004)。

[44] 庄园旗人是辽宁社会史研究和辽宁八旗系统研究的热点。在我们研究盛京内务府其它旗人种类之前，实际上所有的研究辽宁社会史或辽宁八旗系统的学者都集中研究庄园旗人而不是正身旗人。例如由 Guan Jialu, Tong Yonggong (Anoymous 1992)等编纂的有关辽宁小麦庄园的两卷文件，以及许多描述、分析 Guan, Tong (1995)和他们同事数据的文章。也可以参考 Lai 1998，Lee 和 Campbell 2000。

[45] 许多辽宁旗人家属在隶属于户部的“官地”工作。

了我们发现的 36 个此类人群，其中包括 3 个充公庄园人群和两个契约奴役人群[46]。

在辽宁身档案馆和犹他州家谱学会留存的这类充公庄园户口册，标题都是以“（具体产品或制品名称）庄头户口册”或“园头户口册”。这里“庄”说明该庄园是由皇家内务府控制，“庄头”和“园头”则指这些农庄的监管人员。我们从犹他州家谱学会编目的户口册中只找到了 15 部有关庄园人口的，另外还有两部相关的但没有标题，并只深入研究了其中的一部[47]。此外，在这些发现的部分中也包括了各种各样特殊义务旗人人群——他们事实上不属于庄园人口。因而，这些登记才更显重要，其反映了亲属组织结构，庄园组织结构以及个人的人事行为，这些人是那些远多于普通旗人人群和特殊义务旗人人群的剩余辽东地区人口的缩影。

当然，不是所有的庄园人口的情况都一样。总体来说，那些庄园的监管人员以及他们的追随者在庄园人群中有相对较高的地位。他们拥有参加国家科举考试的权利和拥有土地所有权的机会，是因为他们在清朝建立之初把自己的土地贡给清政府以换取旗人身份（Ren 2003; Ding, Guo, Lee, and Campbell 2004）。后来出现的庄园旗人，主要是那些因犯罪而被充公的贵族庄丁，甚至还包括贵族及其家人（Ren 2003）。后面这些庄园旗人的社会地位很低，没有参加科举考试的权利和拥有土地的机会。我们从犹他那里共找出了 6 批这类人群，其中的 3 部：19,24 和 26 包括在 CMGPD-LN 中[48]。在每一例中，人口数据库是依据某一位贵族或官员的姓名（他们的庄园被没收充公）来确定的。在数据库 19 中，阿尔吉善认定户口册,所有的个人是由会计司管理的。其他两个人群数据中的个人则是由正常的旗管理的，尽管他们也属于奴役阶层。

一般而言，登记为庄园旗人的人口同普通旗人相比社会地位较低。如表 20 所示，庄园旗人人口具有不相同的行政管理系统；地区的行政首领不是佐领而是称作“管领”。这一不同便决定了庄园旗人与普通旗人间的不平等。在清代早期，管领是用来管理附属仆人的系统。朝廷不仅禁止管领系统下的官员从事佐领系统中的职位，还禁止这两个系统间家族的通婚（Ding, Guo, Lee, and Campbell 2004）。这种不平等的情况一直持续到 18 世纪中叶。

除了行政管理系统之外，庄园旗人与普通旗人以及特殊义务旗人相比，在内部社会组织结构和地区政权上也有所不同。表 20 中可见，庄园旗人人口在户口册中不是以住户的单位进行组织，而是以父系户组的形式来安排，没有相关的住户信息。此外，尽管每个父系户组都有一位首领（族长），但事实上其并不是真正的族长，而是有行政管理权力的庄头或园头。

还有便是，庄园旗人人口和特殊义务旗人一样，在社会阶层和地理上的流动性很低。庄园旗人人口中没有人拥有官方头衔，除了地区内部的管理者和监管者[49]。

[46] 我们也计划录入 34 卷 Beilu 和 Nanlu zhuangtou deng hukou ce，今后发行。这些户口册数量基本上是正身旗人户口册的 3 到 4 倍，代表了额外 200,000 条观测值。

[47] 这是 1910 Beilu zhuangtou deng hukou ce。

[48] 我们至今辨识出的六个样本是阿尔吉善户口册, the Lu Hongling hukou ce, the Lu Hongquan hukou ce, the Lu Kaiguang hukou ce, 苏拜公户口册, 和五户户口册, 其中 3 个已经被包括进 CMGPD-LN，数据库编号分别为 19, 26 和 24。

[49] 这些地方官员包括族长, 百总, 庄头或园头, 和官领。

5.D 其他奴役人口

除了上述人口以外，盛京皇家内务府下属还管理许多奴役人群，像“抱养”“孤户”“开档（新登记）”“另计（附属）”“苏拉（仆人）”“投从（随从）”。这些奴役人口最初是在清代早期，汉族人通过贡献土地或是作为正身旗人的下属仆人成为八旗系统下的一员。他们中有些人被收养后甚至也获得了正身旗人的身份。然而从 18 世纪中叶开始，由于清政府国库紧张，难以支撑众多旗人的开支，雍正和乾隆皇帝把这些仆人，奴隶还有被旗人所收养的汉族人中的很大一部分清出了八旗系统。尽管这些汉族旗人被清出了八旗，但政府依旧追踪这些人的去向。即对这些人口继续进行分开而独立的人口登记。在 CMGPD-LN 中的两个开档人群，27 和 28 就是典型的例子。

5.E 其他特殊人口户口册

除了上述三类基本人口外，还有一些为特殊的人群和事件而设立的户口册——军队和官员以及特殊义务人口中的军队和官员。这些户口册主要的对象是盛京皇家内务府所附属的多种多样的世袭人群，这些人不是提供专门的劳动产品而是提供专门的服务——通常有士兵，官员及官员下属等等。

除了在户口册中系统地将人口按类划分和管理，政府还为一些特殊事件——比如饥荒后休养——提供了多种多样的户口册。我们在附录 B 中把这些目前已经有所了解的人口类别一一列出。但限于 CMGPD-LN 数据版本中这些类别大多是依据标题来确定的，而非户口册的内容，因此当更多的有效数据和信息被纳入和分析后，这些内容将会有所改进。

6 中国多世代人口数据库—辽宁部分（CMGPD-LN）人口概况

CMGPD-LN 中所涵盖的 29 个行政区域内的人口全面地展现了八旗系统内的社会经济状况。从人群类别看，29 个人群中包含了普通旗人，特殊义务旗人，庄园旗人及其他奴役旗人。并且在每一类人群中，人口的地区分布以及整体的社会地位和成就也各不相同。在本说明的第五部分中，我们主要是对每一类人群所具有的特点进行了总结和分析。首先，我们将对 CMGPD-LN 中所有人群的情况加以概述，接着对每一类人群的特色进行总结讨论。

29 个行政区内的人口，在种族构成、行政类别、人口数量、地理位置方面都存在显著的不同。29 个行政区内有 21 个全部是由汉族旗人构成，其他 8 个——盖州满汉人丁, 八棵树老古洞满汉人丁, 貂皮屯人丁, 王多罗树满洲牲丁, 王多罗树满洲人丁, 苏拜公人丁，棉花盐靛小门人丁和棉花盐靛人丁——由满族、朝鲜族、蒙古族旗人构成。此外，如表 21 中显示，29 个行政区人口数据库的规模，范围从最少的苏拜公人丁中的 879 位个人的 4162 项记录到最多的海城汉军人丁中 20032 位个人的 119186 项记录。这些人口主要居住在辽宁中部和北部，占数据库中全部人口的比例达 68.13%。另外只有王多罗树满洲人丁在有记录的 114 年间（1792-1906）居住地仍保留在最初的地方，其他地区的人口都随着时间的流逝，居住区域扩大分散到其他的地方。

表 21 29 个行政单位的人口规模和地理分布

行政人口		辽宁区域				观察值总计	个人数量
		北部	中部	中南部	南部		
道义屯汉军人丁	频数	24,592	95,019	638	498	120,747	18,884
	百分比	20.37	78.69	0.53	0.41	100.00	
盖州汉军人丁	频数	0	10,647	7,482	24,690	42,819	6,957
	百分比	0.00	24.87	17.47	57.66	100.00	
打蜜汉军人丁	频数	30,320	1,786	0	0	32,106	5,013
	百分比	94.44	5.56	0.00	0.00	100.00	
城内汉军人丁	频数	93	55,585	0	0	55,678	10,054
	百分比	0.17	99.83	0.00	0.00	100.00	
盖州汉军棉丁	频数	0	3,485	4,563	48,000	56,048	10,714
	百分比	0.00	6.22	8.14	85.64	100.00	
牛庄刘二堡喇嘛园人丁	频数	5,395	0	49,178	297	54,870	10,208
	百分比	9.83	0.00	89.63	0.54	100.00	
肥地一面城人丁	频数	65,063	6,302	0	0	71,365	9,543
	百分比	91.17	8.83	0.00	0.00	100.00	
盖州人丁满汉人丁	频数	1,490	595	12,226	38,973	53,284	11,952
	百分比	2.80	1.12	22.94	73.14	100.00	
大甸子汉军人丁	频数	73,727	3,259	0	0	76,986	14,060
	百分比	95.77	4.23	0.00	0.00	100.00	
郭三屯人丁	频数	17,176	16,269	1,636	0	35,081	4,932
	百分比	48.96	46.38	4.66	0.00	100.00	
八棵树老古洞人丁	频数	48,110	650	0	0	48,760	8,247
	百分比	98.67	1.33	0.00	0.00	100.00	
大辛屯人丁	频数	46,451	38,706	1,366	412	86,935	15,748
	百分比	53.43	44.52	1.57	0.47	100.00	
年马岱海寨人丁	频数	52,760	5,464	0	0	58,224	10,746
	百分比	90.62	9.38	0.00	0.00	100.00	
长寨子 汉军人丁	频数	40,987	7,813	0	0	48,800	9,400
	百分比	83.99	16.01	0.00	0.00	100.00	
照化屯人丁	频数	32,063	18,791	0	8	50,862	8,758
	百分比	63.04	36.95	0.00	0.02	100.00	
貂皮屯人丁	频数	80,426	0	86	0	80,512	14,046
	百分比	99.89	0.00	0.11	0.00	100.00	
郎家堡魏家屯人丁	频数	391	12,599	31,465	2,876	47,331	9,092
	百分比	0.83	26.62	66.48	6.08	100.00	
王指挥屯人丁	频数	21,826	28,510	9,994	4	60,334	10,054
	百分比	36.18	47.25	16.56	0.01	100.00	
阿尔吉善人丁	频数	0	8,107	6,991	0	15,098	2,517
	百分比	0.00	53.70	46.30	0.00	100.00	
海城汉军人丁	频数	2,166	711	92,426	23,883	119,186	20,032
	百分比	1.82	0.60	77.55	20.04	100.00	

王多罗树满洲牲丁	频数	9,445	0	49	0	9,494	2,300
	百分比	99.48	0.00	0.52	0.00	100.00	
王多罗树满洲人丁	频数	23,233	0	0	0	23,233	4,899
	百分比	100.00	0.00	0.00	0.00	100.00	
瓦子峪人丁	频数	11,591	59,487	3,832	132	75,042	12,884
	百分比	15.45	79.27	5.11	0.18	100.00	
五户人丁	频数	10,676	11,331	22,783	0	44,790	7,687
	百分比	23.84	25.30	50.87	0.00	100.00	
棉花盐靛人丁	频数	9,382	8,112	32,734	26,510	76,738	13,404
	百分比	12.23	10.57	42.66	34.55	100.00	
苏拜公人丁	频数	557	2,730	875	0	4,162	879
	百分比	13.38	65.59	21.02	0.00	100.00	
开档人丁	频数	4,089	4,527	219	0	8,835	2,717
	百分比	46.28	51.24	2.48	0.00	100.00	
开档投从抱养人丁	频数	10,071	4,741	10,146	5,179	30,137	5,087
	百分比	33.42	15.73	33.67	17.18	100.00	
棉花盐靛小门人丁	频数	1,923	1,643	11,268	11,066	25,900	5,277
	百分比	7.42	6.34	43.51	42.73	100.00	
总计	频数	624,003	406,869	299,957	182,528	1,513,357	266,091
	百分比	41.23	26.89	19.82	12.06	100	

这种人群的逐步分散主要是由于辽东八旗人口整体的地域流动造成的。从表 22 中即可看出，266091 位个人中有 21190 位（7.96%）都曾经被记录从一个地方迁移到另一地方，有 525 位（0.2%）甚至曾经在三个地方定居过。王多罗树满洲牲丁和满洲人丁是流动性最小的人群，而像特殊义务依赖人群开档投从抱养人丁以及从贵族和官员处被充公的庄园旗人人群，如阿尔吉善、五户汉军人丁和苏拜公人丁则是人口流动性最大的群体。

表 22 29 个行政区域人口的地区流动，按照每个个人居住的地区的数量

行政人口	地区数量				
		1	2	3	总计
道义屯汉军人丁	Freq.	14,936	491	18	15,445
	Percent	96.7	3.18	0.12	100
盖州汉军人丁	Freq.	6,253	202	0	6,455
	Percent	96.87	3.13	0	100
打蜜汉军人丁	Freq.	4,006	50	0	4,056
	Percent	98.77	1.23	0	100
城内汉军人丁	Freq.	9,980	74	0	10,054
	Percent	99.26	0.74	0	100
盖州汉军棉丁	Freq.	10,535	178	1	10,714
	Percent	98.33	1.66	0.01	100
牛庄刘二堡喇嘛园人丁	Freq.	9,449	9	0	9,458
	Percent	99.9	0.1	0	100

肥地一面城人丁	Freq.	8,051	123	0	8,174
	Percent	98.5	1.5	0	100
盖州人丁满汉人丁	Freq.	11,182	357	0	11,539
	Percent	96.91	3.09	0	100
大甸子汉军人丁	Freq.	12,041	179	0	12,220
	Percent	98.54	1.46	0	100
郭三屯人丁	Freq.	4,024	903	0	4,927
	Percent	81.67	18.33	0	100
八棵树老古洞人丁	Freq.	7,183	30	0	7,213
	Percent	99.58	0.42	0	100
大辛屯人丁	Freq.	13,609	2,038	94	15,741
	Percent	86.46	12.94	0.6	100
年马岱海寨人丁	Freq.	10,101	278	0	10,379
	Percent	97.32	2.68	0	100
长寨子 汉军人丁	Freq.	8,402	120	0	8,522
	Percent	98.59	1.41	0	100
照化屯人丁	Freq.	6,716	909	0	7,625
	Percent	88.08	11.92	0	100
貂皮屯人丁	Freq.	13,019	11	0	13,030
	Percent	99.92	0.08	0	100
郎家堡魏家人丁	Freq.	8,400	665	18	9,083
	Percent	92.48	7.32	0.2	100
王指挥屯人丁	Freq.	8,920	457	2	9,379
	Percent	95.11	4.87	0.02	100
阿尔吉善人丁	Freq.	2,057	459	0	2,516
	Percent	81.76	18.24	0	100
海城汉军人丁	Freq.	19,216	706	5	19,927
	Percent	96.43	3.54	0.03	100
王多罗树满洲牲丁	Freq.	1,775	4	0	1,779
	Percent	99.78	0.22	0	100
王多罗树满洲人丁	Freq.	4,703	0	0	4,703
	Percent	100	0	0	100
瓦子峪人丁	Freq.	11,544	251	0	11,795
	Percent	97.87	2.13	0	100
五户人丁	Freq.	7,089	396	0	7,485
	Percent	94.71	5.29	0	100
棉花盐靛人丁	Freq.	12,120	448	0	12,568
	Percent	96.44	3.56	0	100
苏拜公人丁	Freq.	745	61	0	806
	Percent	92.43	7.57	0	100
开档人丁	Freq.	1,560	8	0	1,568
	Percent	99.49	0.51	0	100
开档投从抱	Freq.	4,699	81	0	4,780

养人丁	Percent	98.31	1.69	0	100
棉花盐靛小门人丁	Freq.	4,275	79	0	4,354
	Percent	98.19	1.81	0	100
总计	Freq.	236,590	9,567	138	246,295
	Percent	96.06	3.88	0.06	100

29 种人群在国家给予的非农业生产岗位的机会方面存在相当大的区别。表 23 列出的结果显示，29 种人群中有 2.19%的 18-60 岁间男性拥有带俸禄的旗内职位，工匠比例有 0.003%，军人比例 1.41%。拥有职位的个人所占比例在全部 29 种人群中相差很大，像属于庄园旗人人群的阿尔吉善，五户汉军人丁和苏拜公人丁比例为 0 或接近于 0，而像城内汉军人丁的比例则达到 9.77%之多。造成这种差异的部分原因归结于不同类型人群间个人所拥有的权利与机会的差距。举个例子，庄园旗人不被允许拥有官员或军人的位置，而那些本身即登记属于军队或政府的人群中拥有这类岗位的人比例则高很多。同时，只从人群类别角度也无法完全解释这个问题，因为事实上同一人群类别中不同群体间也有很大差异。比如在普通旗人类别中，拥有带俸禄岗位的个人所占比例范围，从最低的海城汉军人丁的 0.51%到最高的郭三屯汉军人丁的 6.49%。一般而言，居住在距离沈阳愈近的人群，拥有非农业生产岗位——工匠、士兵、官员的个人的比例则愈高。

表 23 不同人群数据 18-26 岁男性中有官职、工匠、士兵的数量

行政人口		带俸禄的岗位	工匠	士兵
道义屯户口册-汉军人丁户口册	频数(职位/观察值的总数)	2,198/42,418	607/42,418	1,281/42,418
	占所有观察值的百分比	5.18	1.43	3.02
盖州户口册-汉军人丁户口册	频数	134/13,914	0/13,914	101/13,914
	百分比	0.96	0	0.73
打蜜户口册-汉军人丁户口册	频数	33/11,027	0/11,027	8/11,027
	百分比	0.3	0	0.07
城内户口册-汉军人丁户口册	频数	1,664/17,033	455/17,033	974/17,033
	百分比	9.77	2.67	5.72
盖州棉丁户口册-汉军棉丁户口册	频数	0/19,965	0/19,965	3/19,965
	百分比	0	0	0.02
牛庄刘二堡喇嘛园人丁户口册-汉军人丁户口册-汉军鱼丁户口册	频数	5/19,991	0/19,991	0/19,991
	百分比	0.03	0	0
肥地一面城人丁户口册-汉军人丁户口册	频数	576/25,353	20/25,353	463/25,353
	百分比	2.27	0.08	1.83
盖州人丁户口册-满汉人丁户口册	频数	216/17,739	0/17,739	112/17,739
	百分比	1.22	0	0.63
大甸子户口册-汉军人丁户口	频数	290/24,685	10/24,685	98/24,685

册	百分比	1.17	0.04	0.4
郭三屯人丁户口册-汉军人丁户口册	频数	780/12,019	38/12,019	635/12,019
	百分比	6.49	0.32	5.28
八棵树老古洞人丁户口册-满汉人丁户口册	频数	192/17,354	3/17,354	174/17,354
	百分比	1.11	0.02	1
大辛屯人丁户口册-汉军人丁户口册	频数	838/31,030	239/31,030	523/31,030
	百分比	2.7	0.77	1.69
年马岱海寨人丁户口册-汉军人丁户口册	频数	181/20,626	0/20,626	108/20,626
	百分比	0.88	0	0.52
长寨子户口册 - 汉军人丁户口册	频数	282/15,735	4/15,735	232/15,735
	百分比	1.79	0.03	1.47
照化屯人丁户口册-汉军人丁户口册	频数	591/18,216	133/18,216	401/18,216
	百分比	3.24	0.73	2.2
貂皮屯人丁户口册 - 汉军人丁户口册	频数	214/28,948	0/28,948	202/28,948
	百分比	0.74	0	0.7
郎家堡魏家屯汤岗子海州岳州人丁户口册 -- 汉军人丁户口册	频数	227/16,406	14/16,406	161/16,406
	百分比	1.38	0.09	0.98
王指挥屯人丁户口册 -- 汉军人丁户口册	频数	819/21,389	168/21,389	519/21,389
	百分比	3.83	0.79	2.43
阿尔吉善户口册	频数	0/5,913	0/5,913	0/5,913
	百分比	0	0	0
海州户口册 - 海城汉军人丁户口册	频数	199/38,936	12/38,936	70/38,936
	百分比	0.51	0.03	0.18
王多罗树满洲牲丁户口册	频数	42/3,418	0/3,418	39/3,418
	百分比	1.23	0	1.14
王多罗树满洲人丁户口册	频数	212/7,575	0/7,575	209/7,575
	百分比	2.8	0	2.76
瓦子峪人丁户口册--汉军人丁户口册	频数	1,417/26,417	177/26,417	1,041/26,417
	百分比	5.36	0.67	3.94
五户人丁户口册 -- 汉军人丁户口册	频数	190/15,516	0/15,516	0/15,516
	百分比	1.22	0	0
棉花盐靛人丁户口册—佐领下盐靛庄头壮丁户口册	频数	53/25,978	0/25,978	0/25,978
	百分比	0.2	0	0
苏拜公人丁户口册	频数	3/1,500	0/1,500	0/1,500
	百分比	0.2	0	0
开档户口册	频数	39/2,988	39/2,988	0/2,988
	百分比	1.31	1.31	0
开档投从抱养人丁户口册	频数	8/10,352	0/10,352	0/10,352
	百分比	0.08	0	0
棉花盐靛小门人丁户口册	频数	0/8,876	0/8,876	0/8,876
	百分比	0	0	0
Total	频数	11,403/521,317	1,919/521,317	7,354/521,317

	百分比	2.19	0.003	1.41

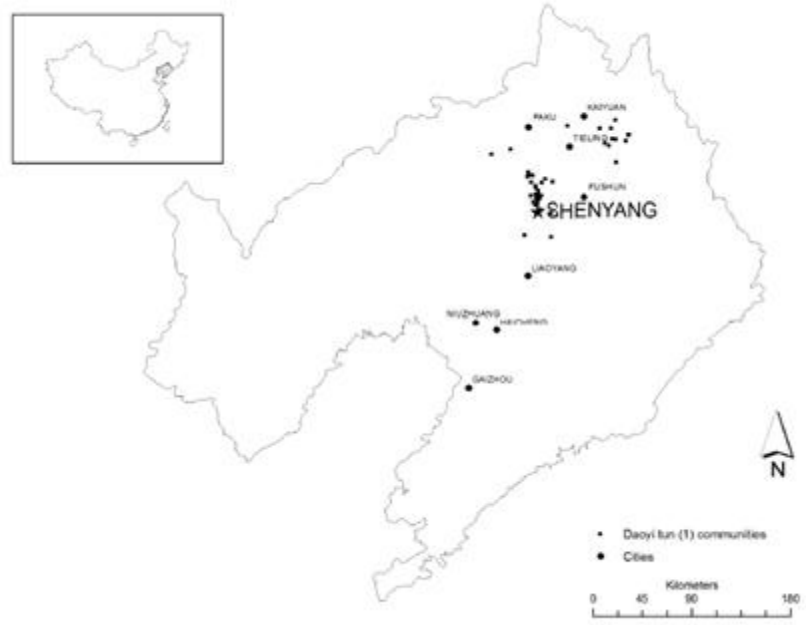
为了让用户对上述 CMGPD-LN 中的 29 种人群中每一人群的历史背景和特色都有更好的理解，我们在下文中将逐一讨论每一类人群的特色。其中包括我们目前为止所能确定的人口的地理位置的地图以及表 21-23 对每一人群相关数据的总结。地理上的流动性依据曾经在两个以上地区生活的个人的比例，分成四个程度：**流动性低**，低于 1%；**流动性一般**，1%-5%之间；**流动性高**，5%-10%之间；**流动性很高**，大于 10%。拥有国家提供岗位的机会则依据人群内拥有国家提供的带俸禄岗位的成年男性的比例，也分为四个程度：**机会少**，低于 0.5%；**机会一般**，0.5-2%之间；**机会多**，2%-5%之间；**机会很多**，大于 5%。这些划分是基于 CMGPD-LN 人口的相对表现，凭经验进行的划分。

6.A道义屯汉军人丁(1)

道义屯汉军人丁数据库是由 1759-1909 年间的 35 卷道义屯汉军人丁户口册——即道义屯汉族旗人户口册录入而来。此数据库共包括 18884 位个人的 120747 项记录。其中道义屯汉军人丁的数量从 1759 年的 1883 人增长至 1909 年的 4178 人[50]。

道义屯位于现今辽宁省省会——沈阳市北部郊区。尽管该户口册是以某一特定地点命名的，但实际上，道义只是人口最初时的行政地理位置。在数据库所涉及的 160 年间，道义屯内的居民共向 88 个其他的村庄进行过迁移，但其大部分、78.69%仍旧居住于辽东中部道义附近的地区（表 21），其他有 20.38%生活在辽宁东部丁家房附近区域——这一地区曾属于铁岭，现在归为法库地区。丁家房的多部宗谱记载了从道义到丁家房的迁徙事件，并且从道义户口册可以将迁徙开始时间定在 1792 年。剩下还有一少部分居民位于辽宁南部和中南部，主要是辽阳地区附近。

[50] 在 1997 年的著作中，James Lee 和 Cameron Campbell 基于这些人口数据，研究了道义地区人口的社会人口史(Lee and Campbell 1997)。



地图 3 道义地区人口

表 24 道义屯汉军旗特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	有
荣誉头衔的数量	10
是否可以参加科举考试	是
地理流动性	高
带俸禄的职位	非常高

从名称“汉军人丁户口册”可知，道义屯户口册所载人群为普通旗人。另外，由于道义屯人口主要居住于沈阳附近——沈阳是清未统一全国前的首都，也是最初皇家宫殿所在地，故其中的成年男性拥有带俸禄的职位的比例比较高。如表 23 显示，道义屯人口中有 5.18%的 18-60 岁男性拥有带俸禄职位。此比例超过 CMGPD-LN 中的平均水平 2.19%两倍还多。

类似地，道义屯 18-60 岁男性中，拥有工匠或军人身份的比例也比整个数据库中同样年

龄段的平均水平高很多。此外，工匠从事的行业也很多样化，包括弓匠、锞子匠、染匠、花儿匠、裁缝、皮匠等等。同其他普通旗人人群相比较，道义屯人口有较多机会从事国家提供的岗位，仅次于郭三屯和城内两地的在职男性的比例。

6.B盖州汉军人丁(2)

盖州汉军人丁数据库是由 1759-1909 年间的 35 卷盖州汉军人丁户口册——即盖州汉族旗人户口册录入而来。此数据库共包括 6957 位个人的 42819 项记录。这一数据库的人口规模从 1759 年的 793 人增长至 1909 年的 2565 人。

地图 4 盖州汉军旗人口

盖州位于辽东半岛西北部，是辽宁南部的多山沿海区域。盖州的东部和东南部主要为山区，只有北部和西部地势相对平坦。盖州户口册中的人口来自 47 个村庄。

表 25 盖州汉军旗特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长

军事长官	有
工匠	无
地方官员	无
荣誉头衔的数量	3
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	中等

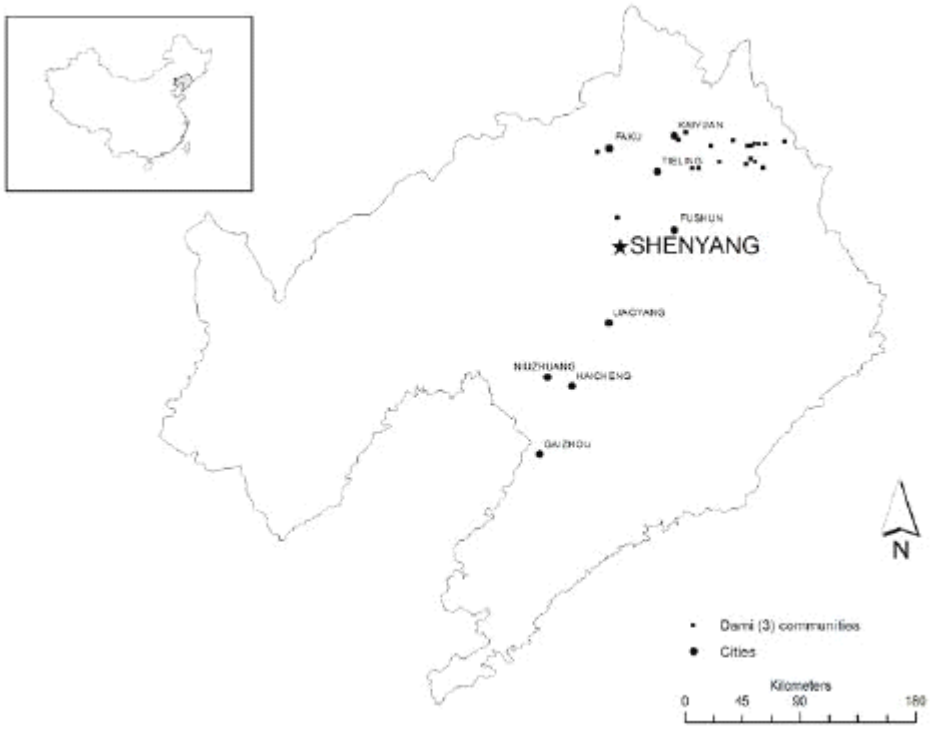
和道义屯汉军人口类似，盖州汉军人口也主要分散在辽宁省内的三个地区：如表 21 所示，57.66%的相关记录分布于辽宁南部，24.87%的记录分布于辽宁中部，17.47%的记录分布于辽宁中南部。对不同年代人口的地理分布做一仔细研究可以发现，1789 年之前，盖州汉军人口主要居住于辽宁南部，自 1789 年开始户口册中显示有人口开始定居在辽宁中部和中南部。

盖州汉军人口也属于普通旗人人口，但同道义屯汉军人口相比较，18-60 岁男性中从事国家提供岗位的人的比例要低很多，只有约 0.96%的 18-60 岁男性拥有带俸禄岗位（表 23）。没有一个个人有工匠身份，只有一少部分人的记录显示为军人，而这部分军人的职位也普遍较低[51]。

6.C打蜜汉军人丁(3)

同道义屯和盖州汉军人丁户口册中以某一地名作为名称不同，打蜜汉军人丁户口册则是以该人口向皇家内务府所提供的劳动服务的种类——此处即养蜂——来命名的。打蜜汉军人丁数据库是由 1759-1909 年间的 32 卷打蜜汉军人丁户口册录入而来的。此数据库共包括 5013 位个人的 32106 项记录。

[51] 这些低职位包括执事人、笔帖式和库使。



地图 5 打蜜汉军旗村庄分布

打蜜汉军户口册中的个人分布于 49 个村庄。然而，其中绝大多数（依据表 21 为 94.44%）都生活在辽宁北部。那里主要是山区，在 18 世纪时，人烟稀少，所以特别适合养蜂产蜜。鉴于其特殊的劳动义务，这些采蜜人的居所流动性比道义和盖州的农民低得多，只有 5.56%的记录显示其出现在其他地区[52]。表 22 显示，道义户口册中有 5.93%的个人、盖州户口册中有 7.49%的个人曾在超过两个地区居住，而打蜜汉军人丁户口册所载人口中只有 2.25%的个人曾在辽宁北部之外的地区生活。

表 26 打蜜汉军旗特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	窝棚头
行政长官	族长
军事长官	有
工匠	无
地方官员	无
荣誉头衔的数量	0

[52] 养蜂产蜜的一个主要用途就是皇家陵园的祭祀。清朝早期皇帝的陵园都位于辽宁中部和北部。

是否可以参加科举考试	否
地理流动性	中等
带俸禄的职位	低

打蜜汉军人丁户口册中所载人口的规模没有像前面两种人口那样增长很快，相反在一个半世纪以来波动甚小。人口从 1759 年的 1060 人，到 1792 年跌至 1000 人以内，至 1812 年只有 849 人，1849 年 726 人，之后直到 1888 年维持在 750 人左右。到 1903 年所登记人口才增长到 1029 人，最后到 1909 年为 1137 人。

打蜜汉军人丁主要从事采蜜，成年男性被归为蜜丁。任一年中，蜜丁的人数大约都有 300 左右。每位蜜丁每年需要上缴两桶半的蜂蜜（约合 56 磅）给皇家内务府（Ding, Guo, Lee, and Campbell 2004）。一般情况下，一个天然蜂窝一年可以产约 50 磅的蜂蜜，故这些蜜丁同皇家内务府下属的其他生产人口相比，工作任务还是相对轻的。

同道义和盖州的人口相比，打蜜汉军人口从事国家提供岗位的机会很有限。表 23 中显示，只有 0.3%的打蜜汉军中 18-60 岁男性有带俸禄职位。至今我们还不清楚这一结果有多大程度源于其从事的特定产业——他们主要居住在远离城市的偏远地区。

然而，自 18 世纪晚期开始，蜜丁可以根据其原本应该向向皇家内务府提供的蜂蜜数量、以每桶蜂蜜 0.5 两白银的价格支付货币，以免去这项劳动义务。因此，有些蜜丁变为了普通的农民（Ding, Guo, Lee, and Campbell 2004）。

6.D城内汉军人丁(4)

城内汉军人丁数据库是由 27 卷 1789-1909 年间的城内汉军人丁户口册录入而来。这里的“城”是指沈阳。此数据库共记录了 10054 位个人的 55678 项记录，这其间，其所属人口几乎翻了一番，从 1789 年的 1865 人增长至 1909 年的 3270 人。城内户口册中所记载人口大多属于城市驻防军人和官员的家庭。

表 27 城内汉军人口特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	有

荣誉头衔的数量	29
是否可以参加科举考试	是
地理流动性	低
带俸禄的职位	非常高

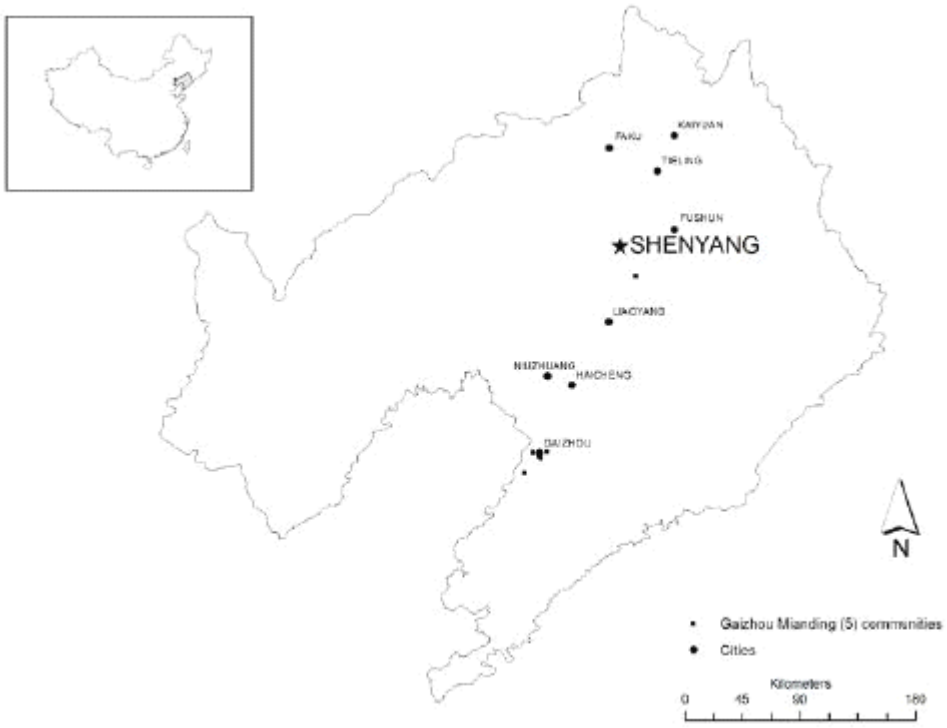
从 CMGPD-LN 所覆盖的整个时间段来看，城内汉军人口几乎没有大的地理上迁移，基本都集中在沈阳周围。如表 27 显示，城内汉军人口户口册所载 99.83%的人口都居住在辽宁中部，并且，99.26%的人口从未迁移出这一区域，迁出的个例也都是发生在很晚期。如果按年份细致观察可以发现，1789-1906 年间，所有在户口册上记载的人口均显示居住在沈阳，而只在最后一次登记——1909 年——显示有少部分人迁到了辽宁北部。

在全部 29 种人群中，城内汉军人口普遍的社会地位是最高的。表 23 中就可见，其 18-60 岁男性中有 9.77%都拥有带俸禄的职位，包括像军队士官、教头首领等等。在该人口中，成年男性中工匠和军人的比例也很高。因此，同其他人口相比，城内汉军人口有着非常独特的职业状况。

也许是因为城内汉军人口多属于上层社会成员，其户口册在记录个人女性子嗣方面也是全部 29 个人群中做的最好的。尽管，忽略 16 岁以下的女性子嗣信息是皇家内务府所管理的所有人口统计册的通病，但相比较而言，城内户口册这方面的记录更多些。统计发现，在整个 CMGPD-LN 中，1-5 岁人群的性别比例为 699:100，而城内人口的相应比例仅为 236:100，由此可见一斑。

6.E 盖州汉军棉丁(5)

盖州汉军棉丁数据库是由 1789-1909 年间的 25 卷盖州汉军棉丁户口册录入而来，在此期间，所载人口从 1789 年的 1017 人增长至 1909 年的 5413 人。这一名称很明显的告诉我们：人口的居住地——盖州，专门的义务——生产提供棉花。



地图 6 盖州汉军棉丁村庄分布

表 28 盖州汉军棉丁人口特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	庄头
行政长官	族长
军事长官	无
工匠	无
地方官员	有
荣誉头衔的数量	13
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	无

盖州棉丁人口有四方面特色值得注意：第一，人口居住极为集中，10714 位个人的 56048 项记录都显示位于仅仅七个村庄，其中六个位于盖州内，另一个位于沈阳南部。

第二，其居民地理上的流动性很有限。表 21 中显示，盖州汉军棉丁人口户口册所载人口绝大部分生活在辽宁南部，比例占 85.64%；只有 8.14%的人口位于辽宁中南部。记录显

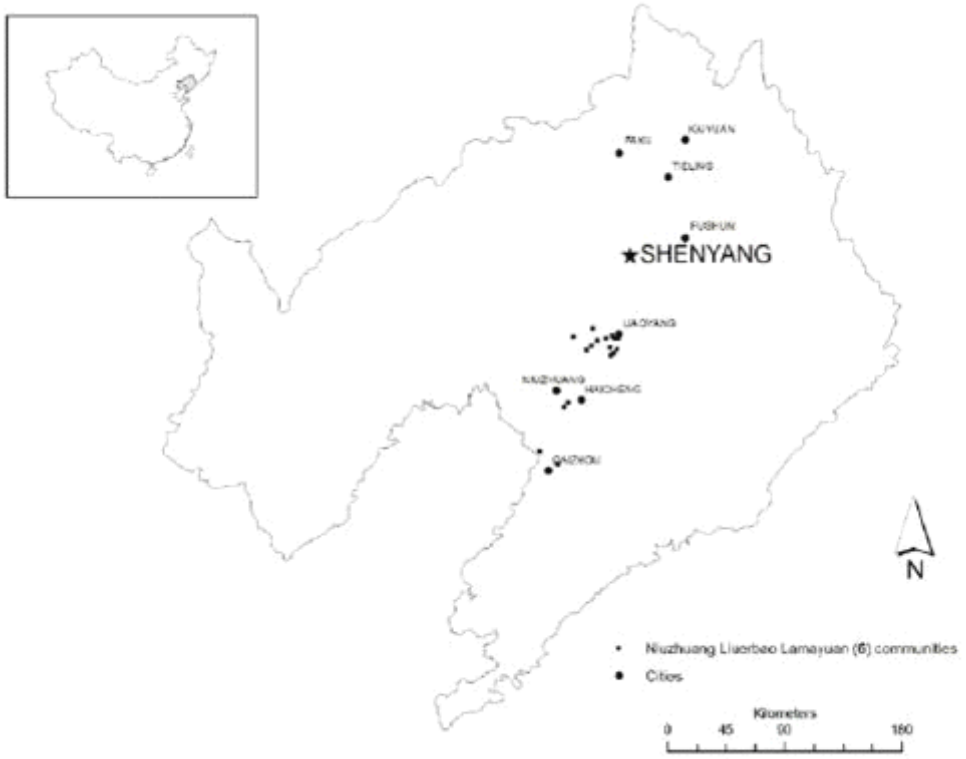
示，1843 年，大约占比例为 6.22%的人口迁居到沈阳南部。除了这次在 1841-1843 年间的相对大的迁移之外，再没有其他大的迁移事件。因此，其人口中只有 1.66%曾在两个以上地区生活过。

第三，和那些打蜜人丁类似，盖州棉丁人口中没有一个人拥有国家制定的非农业生产职位。表 23 中可见，其人口中 18-60 岁男性中无一拥有带俸禄职位。事实上，也有一少部分棉农通过纳贡而换来一些荣誉性头衔，这就显示了，虽然这部分棉丁人口整体的社会地位较低，但其中也有一些比较富裕的家庭。

第四，记录中对个人的女性子嗣的忽略在此数据库中尤其明显。年纪在 16 岁以下的女孩儿极少被登记。比如，年龄在 1-5 岁的人口中，共有 2462 个男孩儿的记录，却只有 48 个女孩儿的记录。如此明显的对女性子嗣的忽略，很有可能是由人口自身较底层的社会经济地位决定的。

6.F 牛庄刘二堡汉军人丁(6)

牛庄刘二堡汉军人丁数据库是由 1777-1906 年间的 25 卷牛庄刘二堡汉军人丁户口册录入而来。该户口册的其他称谓有“牛庄刘二堡喇嘛园人丁户口册”和“牛庄刘二堡汉军鱼丁户口册”。这里的“牛庄”、“刘二堡”、“喇嘛园”都是地名。牛庄和刘二堡位于辽宁中南部，喇嘛园位于辽宁北部。“鱼丁”意味着这一人口的主要义务是向皇家内务府提供鱼产品。依据 Ding, Guo, Lee, and Campbell (2004)的研究，生活在牛庄和刘二堡的主要是渔民，而那些生活在喇嘛园的则属于“坟丁”，即负责看守皇家坟墓。



地图 7 牛庄刘二堡汉军村庄分布

表 29 牛庄刘二堡汉军人口特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	鱼千总
行政长官	族长
军事长官	无
工匠	无
地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	是
地理流动性	低
带俸禄的职位	无

该数据库共记载了来自于辽宁中南部、南部和北部的 35 个村庄的 10208 位个人的 54870 项记录。其人口规模也发生过波动。所登记人口 1777 年开始时为 1316 人，到 1858 年增长至 2970 人。1867 年开始，所有生活在喇嘛园的人口的登记被撤出了牛庄刘二堡户口册，因

此人口规模回落到 1341 人，这些剩下的渔民人口到 1906 年增长至 1941 人。

同其他提供专门义务的人群类似，牛庄刘二堡鱼丁人口的流动性也很有限[53]。绝大部分（89.63%）人口居住在辽宁中南部的牛庄和刘二堡两地（表 21）。只有一小部分（9.83%）人口居住在辽宁北部的喇嘛园，主要负责看守皇家坟墓。其中在 1825 年，还有几户迁到了辽宁南部的牛庄——辽东半岛西部沿海，靠近辽河河口——这一地区是中国重要的东北部港口。自 1860 年始，牛庄成为东北地区第一个贸易港口。刘二堡同样也位于肥沃的辽河冲积平原上。

与其他特殊义务人群一样，牛庄刘二堡汉军人丁中男性普遍政治地位偏低。表 23 显示，在整个时期里，该人口中 18-60 岁男性拥有国家制定的带俸禄岗位的只有 5 位。没有任何人被登记作为工匠和军人。作为皇家内务府下的鱼丁，他们每年需上缴 556 磅的鱼产品（Ding, Guo, Lee, and Campbell 2004）。到后来，自十八世纪晚起开始，他们被允许通过支付银两的方式来兑现捕鱼的义务。

6.G肥地一面城汉军人丁(7)

所有 29 种人群的人口数据库中，肥地一面城汉军人丁数据库是人口记录最为全面的一个。不仅是由最多卷数（39）的户口册录入而来，并且其中有 32 卷都是三年期连续无缺漏的，是所有数据库中最长的。该数据库共包括了生活在 1756-1909 年间的 9543 位个人的 71365 次记录，几乎是每位个人有八次记录，比其他数据库中每位个人 5-6 次记录多出了近三分之一。这些人口主要居住在辽宁中部和北部（开原和铁岭）等地的 30 个村庄，总人口从 1756 年的 1196 人增长至 1909 年的 2533 人。

[53] 由于这个系列的户口册在 1789 年刚刚开始记录旗界，居住地的地理信息在之前并不明确。因为登记格式的变化，这支人口地理分布与年份的列联表显示了在 1789 年人口从辽宁北部和中南部有大规模的迁移。这个迁移假象导致了这些人口明显的地理流动，如表 21 和 22 所示。然而这次迁移很有可能是旗界的分类造成的。

地图 8 肥地一面城汉军人丁村庄分布

表 30 肥地一面城汉军人口特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	有
荣誉头衔的数量	5
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	高

肥地一面城汉军户口册中所载人口主要是由普通旗人和特殊义务旗人组成。其中普通旗人占大多数，有些作为牧丁。其他一些作为仆人的称作“夫丁”，仪式参与人员称作“礼丁”。剩下一些是八旗所属庄园或农场的下层管理者。如表 21 显示的，该人口中绝大部分（91.17%）

都生活在开原地区肥地和一面城两地。只有 8.83%的记录显示人口居住在辽宁中部。此人口的地理流动性一般，在两个地区内生活过的人口比例为 3.56%（表 22）。

鉴于其人口组成主要为普通旗人，肥地一面城人口户口册所载人口整体从事非农业岗位的机会要比那些采蜜、拾棉、打渔的特殊义务旗人要多很多。表 23 显示，18-60 岁间男性的记录中有 2.27%拥有带俸禄岗位，有 1.83%被登记为士兵。虽然这一比例达不到道义汉军人口的水平，但却是远远高于其他的特殊义务人群。此外，男性记录中显示作为工匠的比例并不高，只有 0.08%，主要为生产弓箭的匠人。

6.H盖州满汉人丁(8)

盖州满汉人丁是由 1753-1909 年间的 21 卷盖州满汉人丁户口册录入而来。数据库共有 11952 位个人的 53284 次记录。这些人口是生活在辽宁南部盖州，辽宁北部开原和铁岭，辽宁中部沈阳，辽宁中南部岫岩、辽阳、牛庄等地的 71 个村庄。其人口规模从 1753 年的 1055 人增长至 1909 年的 5128 人。

地图 9 盖州满汉人丁村庄分布

表 31 盖州满汉人丁特征

人口地位	常规旗人

民族	汉、高丽满族
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	无
地方官员	有
荣誉头衔的数量	17
是否可以参加科举考试	是
地理流动性	高
带俸禄的职位	中等

在所有数据库中，盖州满汉人丁数据库的最显著特色是其特殊的民族组成。因为，皇家内务府所辖的大部分人口都为汉族旗人——即汉军，而盖州满汉人口却包括有汉族和满族。事实上，这些登记为满族的人口并不是普通满族人，而是被称作“高丽满洲”。这是所有数据库中除编号 25 的“棉花烟点庄头庄丁”数据库以外唯一记录了朝鲜满族的数据库。在 1864 年以前，当这部分户口册还称作“盖州人丁户口册”时，其中只有很少几户朝鲜满族家庭。而在 1867 年左右，便增加了 60 户。在此之后，朝鲜满族户数基本维持在这个数目，最多时达到 80 户，直到 1903 年一直没有大的缩减。朝鲜满族人口的出现从一个角度显示了八旗人口的种族的潜在多样性。然而，我们对朝鲜满族人口的了解还极为有限。

盖州满汉人口属于普通旗人。如表 21 显示，尽管盖州满汉人口似乎遍布辽宁，实际上大多数人口都生活在辽宁南部的盖州周边区域。此人口的地理流动性一般，有 6.72%的记录显示人口曾在两个地区生活。表 23 显示，1.22%的 18-60 岁男性记录显示其拥有带俸禄的职位，但没有任何作为工匠的例子。

6.1 大甸子汉军人丁(9)

大甸子汉军人丁数据库是由 1756-1906 年间的 27 卷大甸子汉军人丁户口册录入而来。该数据库共记录了 14060 位个人的 76986 次记录。其人口规模从 1756 年的 1828 人增长至 1906 年的 5603 人。

地图 10 大甸子汉军村庄分布

表 32 大甸子汉军村庄人口特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	有
荣誉头衔的数量	11
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	中等

大甸子人口户口册中所载人口属于普通旗人。这些人口尤其集中（95.77%，表 21）在辽宁北部铁岭周边一带，仅有 4.23%的人口生活在辽宁中部。该人口的地理流动性较低，表 22 中显示，只有 1.76%的个人曾经在两个地方生活。

作为居住在远离政治中心的普通旗人人群，大甸子人口中男性成员能享受到一定的政治和社会自由。表 23 可见，1.17%的 18-60 岁男性的记录显示其拥有带俸禄岗位。不过，工匠依然很少，只有 10 次 18-60 岁男性的记录显示其作为工匠（表 23）。

6.J 郭三屯汉军人丁(10)

郭三屯汉军人丁数据库是由 1774-1909 年间的 34 卷郭三屯汉军人丁户口册录入而来。此数据库也属于较为完整的数据库之一，共记录了 4932 位个人的 35081 次记录——平均每位个人有超过 7 次记录，接近肥地和郭三屯。这些人主要居住在辽宁北部、中部、中南部的 21 个村庄。人口规模从 1774 年的 579 人增长至 1909 年的 1335 人。

地图 11 郭三屯汉军人丁村庄分布

表 33 郭三屯汉军人丁特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长

军事长官	有
工匠	有
地方官员	无
荣誉头衔的数量	1
是否可以参加科举考试	是
地理流动性	非常高
带俸禄的职位	非常高

尽管郭三屯汉军人口数量不大，但其地理分布却是非常分散。其中有约一半人口居住在辽宁中部沈阳的东北部郊区——这也可能是此人口的发源地，其余一半人口生活在辽宁北部开原的东昌或是临近的铁岭的商家簸子和拉塔湖等地。鉴于其广泛的分布，该人口的地理流动性自然较高。表 22 中，18.57%的人口曾经在两个地方生活过。

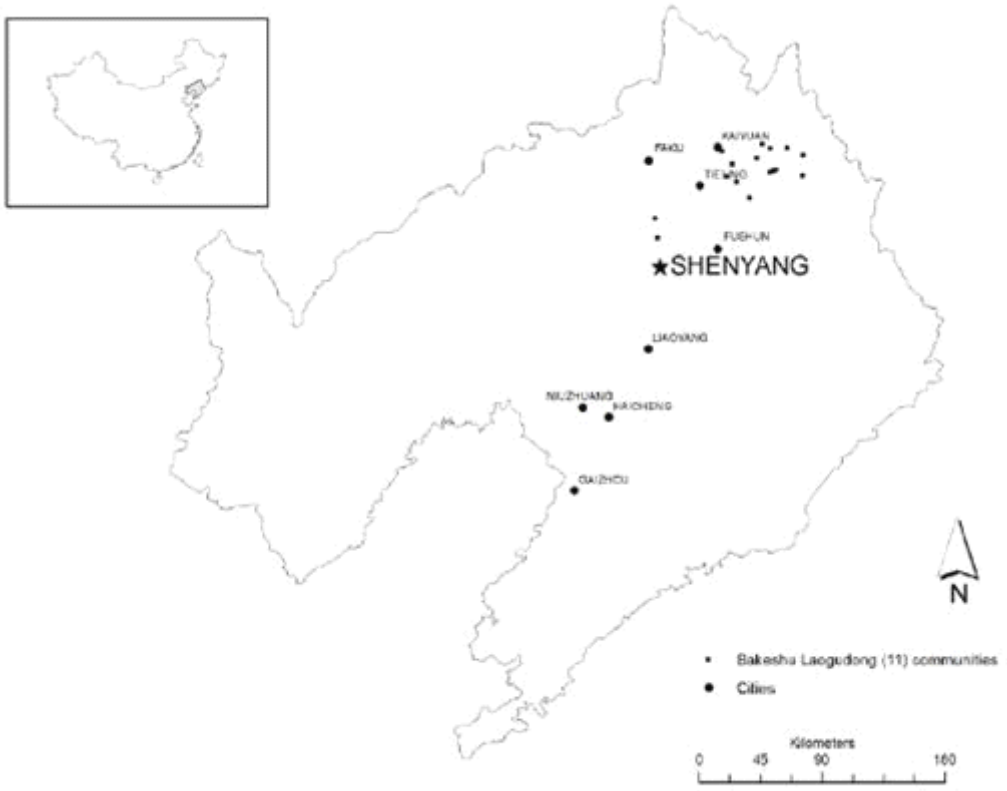
郭三屯旗人人口中，成年男性中拥有国家提供的带俸禄岗位的比例非常高。表 23 中可见，6.49%的 18-60 岁男性的记录显示其拥有带俸禄岗位，5.28%记录显示其作为军人。这些比例几乎可以和城内那些上层社会旗人相媲美，超过了道义汉军人口的水平。特别是，那些生活在沈阳东北部郊区的郭三屯汉军人口，男性中有极高的比例拥有职位。当我们细致分析那些在职人口的地理分布时，发现有 70%都居住在辽宁中部。如果我们仅仅看那些生活在辽宁中部的人口，男性中拥有非农业带俸禄岗位的比例达到了 10.18%，其中大多数为军人，此比例甚至超过了城内人口。因此，郭三屯人口中这些生活在辽宁中部的人群中很可能是有一些军队家庭。

同时，郭三屯人口同道义人口和城内人口相比，作为工匠的个人的比例要少很多。表 23 中可见，只有 0.32%的 18-60 岁男性的记录显示其作为工匠，比道义人口的 1.43%和城内人口的 2.67%低很多。此外，拥有荣誉头衔的个人比例也比其他一些人口要少，比如像海城，有 47 例荣誉头衔拥有者。这一情况很可能反映出郭三屯人口虽然有证据显示其人口有较好的发展和被雇佣机会，但其仍然相对贫困的事实。

郭三屯旗人户口册对于女性子嗣的登记则相对完整一些。记录中 1-5 岁个人的性别比例为 340:100。

6.K八棵树老古洞满汉人丁(11)

八棵树老古洞满汉人丁数据库是由 1759-1909 年间的 32 卷八棵树老古洞满汉人丁户口册录入而来。数据库共包括了 8237 位个人的 48760 次记录，这些人口来自于辽宁北部和中部的 31 个村庄。其人口规模从 1759 年的 930 人增长至 1909 年的 2354 人。



地图 12 八棵树老古洞满汉人丁村庄分布

表 34 八棵树老古洞满汉人丁特点

人口地位	常规旗人
民族	汉、满族
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	有
荣誉头衔的数量	0
是否可以参加科举考试	是
地理流动性	低
带俸禄的职位	中等

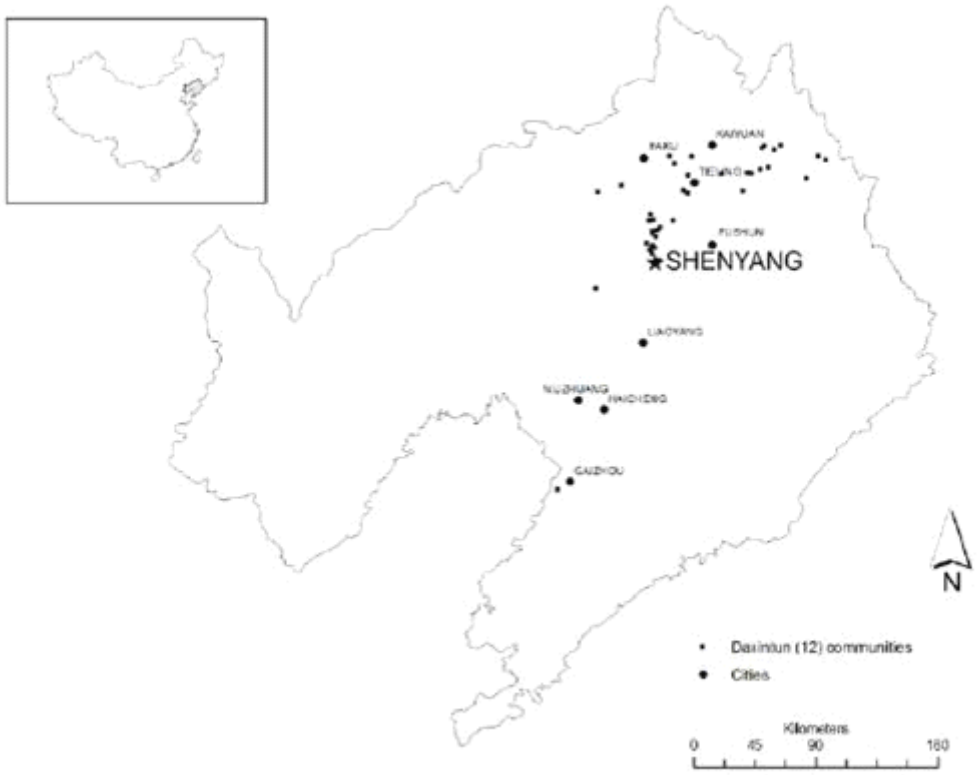
该人口中绝大多数（98.67%）登记都显示其居住在辽宁北部的铁岭和开原两地（表 21），剩下极少数居住在辽宁中部。除此之外，该人口的地理流动性几乎不存在，只有 0.45%的个人曾在两个地区生活过。

鉴于八棵树老古洞满汉旗人人口属于普通旗人，故其中的男性个人是允许从事国家制定的非农岗位的，但就这一人口看，他们的机会属一般水平。表 23 中，仅有 1.1%的 18-60 岁男性记录显示其拥有带俸禄的非农岗位。此外，和郭三屯人口类似，人口中作为工匠的个人比例很低。

与盖州满汉人丁数据库一样，八棵树老古洞满汉人丁数据库也因其不仅有汉族成员、还有满族旗人人口而特殊，尽管事实上只有很少数量的满族居民，而他们也仅仅在 19 世纪中叶才加入进来。1852 年，两个户组（族）——包括 16 户居民开始出现在户口册中，并且其中每一户都有对应的注释——“满洲”——来标明其身份。

6.L 大辛屯汉军人丁(12)

大辛屯汉军人丁数据库是由 1749-1909 年间的 29 卷大辛屯汉军人丁户口册录入而来的。数据库共包括 15748 位个人的 86935 次记录。和肥地与郭三屯数据库一样，大辛屯数据库也是最为完整的数据库之一，每位个人约有接近 8 次的记录留存。居住地主要位于四个地区的 77 个村庄（辽宁北部、中部、中南部、南部）。人口规模从 1749 年的 1218 人增长至 1909 年的 4510 人。



地图 13 大辛屯汉军人丁村庄分布

表 35 大辛屯汉军人丁特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	有
荣誉头衔的数量	12
是否可以参加科举考试	是
地理流动性	非常高
带俸禄的职位	高

大辛屯汉军人口主要居住于辽宁北部和中部等区域：开原、铁岭和沈阳北部郊区。开原和铁岭两地（53.43%）与沈阳北郊（44.52%）人口分布接近。剩下很少一部分生活在辽宁中南部和南部。居住区域的广泛分布即显现了该人口相对高的地理流动性。表 22 显示，有 2038 位个人，占比例为 12.94%都曾在两个地区生活过，还有 94 位个人甚至在三个地方生活过。

作为普通旗人群体[54]，大辛屯汉军人口也能通过从事非农业岗位获得一定的社会阶层提升。如表 23，2.7%的 18-60 岁男性记录显示其从事非农岗位，有国家提供的俸禄。还有 0.77%的记录显示个人有工匠的身份，1.69%记录显示个人作为军人。然而，从事这些岗位的机会在该人口生活的四个地区间分布并不均衡。深入研究可以发现，这些拥有非农岗位的个人都集中于生活在辽宁中部和中南部的群体中。生活在辽宁中部的人口中，有 4.5%的男性记录显示其拥有非农岗位。此外，尽管那些生活在辽宁南部的人口数量很少，但他们中拥有带俸禄岗位，特别是作为工匠的个人却更常见。这一情况即反映，尽管都属于同一户口册下的人口，居住区域的不同也会造成个人在改变其社会地位机会上的很大差异。

6.M 年马岱海寨汉军人丁(13)

年马岱海寨汉军人丁数据库是由 1749-1909 年间的 31 卷年马岱海寨汉军人丁户口册录入而来。数据库共包含 10746 位个人的 58224 次记录，这些人口主要生活在辽宁北部和中部的 54 个村庄。尽管整体上看，其人口规模从 1749 年的 1378 人增长到 1909 年的 2496 人，约翻了一番。实际上，在头一个世纪里，其人口规模的变化趋势并不明了：1749-1864 年间，人口一直在 1300-2000 人左右波动。直到 1867 年，该人口被重新划分为皇家内务府下辖人口后，人口才开始持续的增长。

[54] 虽然大部分大辛屯汉军人口都住在正规国营农场，在早期户口册中，也有一些特殊地位的个人如开档，投从和抱养，这些家庭分布分散，在后来的户口册中没有出现。

地图 14 马岱海寨汉军人丁村庄分布

表 36 马岱海寨汉军人丁特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	窝棚头、养蜂百总
行政长官	族长
军事长官	有
工匠	无
地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	中等

不同于那些普通旗人人口，年马岱海寨人口需要负担特殊的义务，遵守一定的规定。皇家内务府把人口中所有男性分为两类：棉丁（采棉）和蜂丁（养蜂）——即棉丁需要采集棉花，蜂丁则需养蜂产蜜。该人口主要（90.62%）居住在辽东北部地区：棉丁户多居住在开

原的年马地区，蜂丁户则一般居住在铁岭的岱海寨等地。另外还有一部分（9.38%）的蜂丁户居住在辽宁中部、抚顺西北地区（表 21）。这些生活在铁岭和开原等地的人口虽然管理划分上不同（棉丁与蜂丁），但居住区域上较接近并连续。

鉴于年马岱海寨人口的身份为特殊义务旗人，所以这些人居住区域的地理流动性也很一般。表 22 显示，仅有 3.67%的人口曾经被记录在两个地区生活过。类似的，这一人口中个人社会地位的活动性也较低，拥有非农业带俸禄岗位的比例很低。如表 23 所示，仅仅有 0.88%的 18-60 岁男性的记录显示其拥有非农带俸禄岗位。没有人被登记作为工匠，也只有 0.52%的男性记录显示其作为军人。

6.N长寨子汉军人丁(14)

长寨子汉军人丁是由 1768-1909 年间的 27 卷长寨子汉军人丁户口册录入而来。数据库共包括 9400 位个人的 48800 次记录，这些人主要居住在辽东北部地区。像其他生活在辽东北部地区人口一样，其人口规模增长缓慢，从 1768 年的 1411 人增长至 1909 年的 2344 人。

地图 15 长寨子汉军人丁村庄分布

表 37 长寨子汉军人丁特征

人口地位	常规旗人

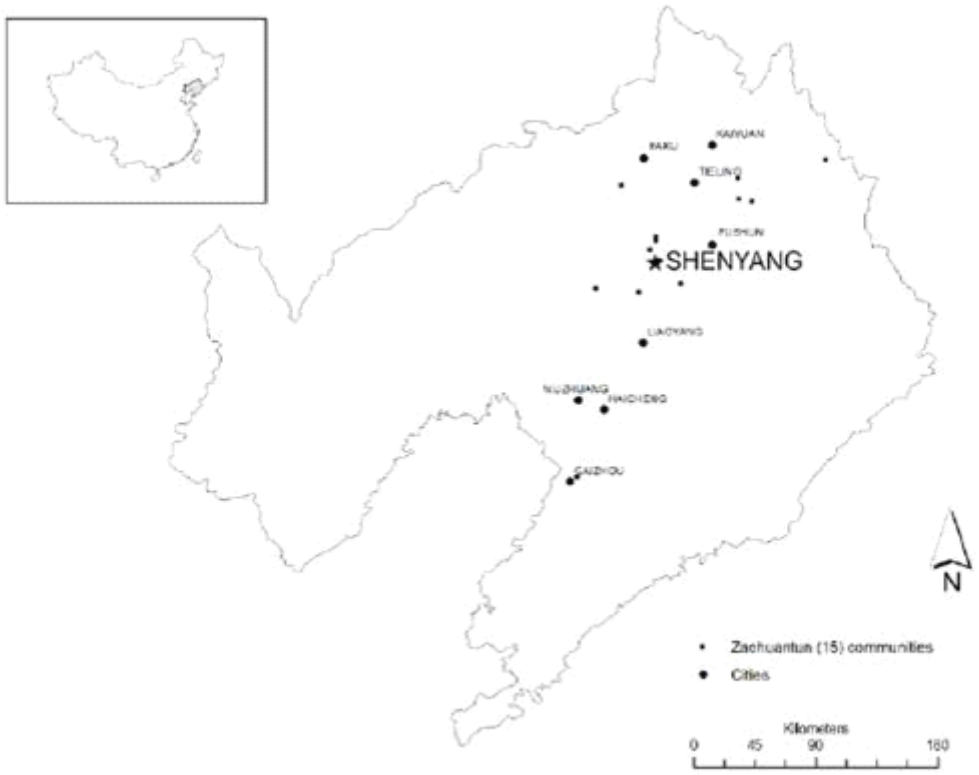
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	有
荣誉头衔的数量	0
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	中等

长寨子汉军人口主要居住在辽北的开原，另外有 16%的人口居住在辽中的 xingjing 和沈阳等地（表 21）。虽然，该人口属于普通旗人，并且大部分人口居住在辽东中部。和年马岱海寨人口相比，只有 2.7%的个人曾经在辽东的各地区间迁移过，这一比例甚至比许多特殊义务旗人人口还要低。

鉴于该人口属于普通旗人，自然人口中男性拥有非农带薪岗位的比例要超过那些属于特殊义务旗人的人口。如表 23 所示，1.79%的 18-60 岁男性记录显示其目前拥有非农带薪岗位，1.47%显示其作为军人。但此人口中，还没有发现有个人作为工匠。

6.0照化屯汉军人丁(15)

照化屯汉军人丁数据库是由 1774-1909 年间的 26 卷照化屯汉军人丁户口册录入而来。数据库共包括 8758 位个人的 50682 次记录，这些人主要居住在辽东北部和中部。和这一地区其他人口类似，登记人口规模从 1768 年的 1411 人增长至 1909 年的 2344 人。



地图 16 照化屯汉军人丁村庄分布

表 38 照化屯汉军人丁特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	无
荣誉头衔的数量	8
是否可以参加科举考试	是
地理流动性	非常高
带俸禄的职位	高

照化屯汉军人口有两大特色。其一，尽管该人口登记的起源地——照化屯——位于辽中沈阳的西北郊区，但事实上，其大多数人口则生活在辽宁北部。如表 21 所示，63%的记录显示个人居住在辽北铁岭地区。其余的，37%，生活在沈阳市郊。作为普通旗人，这一人口

居住区域有一定的流动性。表 22 中，在 1774-1909 年间，有 13.8%的个人曾经在辽宁两个地区生活过。

第二，作为普通旗人，照化屯旗人人口也有一定的机会来从事国家提供的非农带薪岗位。表 23 中，有 3.24%的 18-60 岁男性记录显示其有非农带薪岗位，0.73%有工匠身份，另外有 2.2%为军人。这些男性中拥有非农带薪岗位的比例比 CMGPD-LN 中许多其他人群都要高。再一点即是，不同于郭三屯和大辛屯人口中，拥有非农带薪岗位的男性多集中在生活于辽中地区的人口，照化屯人口中，这一人群在辽中和辽北两地是均匀分布的。

从 1903 年开始，旗人人口归属管理发生过一次相对大的调整。旗人一般应居住在旗地上，而开始出现一部分居民报告他们的新居所不属于任一旗单位。到 1906 年户口册中显示，照化屯人口中有 52 族宣称不再耕种八旗土地并在其上生活。这种旗人身份的改变一方面可能是由于登记中的漏误，另一方面很可能反映的是真实的改变——或许是即将到来的辛亥革命的前兆（Enatsu 2004; Zheng 1981）。此外，在同一时期，还有一户生活在辽宁南部的家庭增加到该户口册中。

6.P貂皮屯汉军人丁(16)

貂皮屯汉军人丁数据库是由 1768-1909 年间的 30 卷貂皮屯汉军人丁户口册录入而来。数据库共包括 14046 位个人的 80152 次记录，这些人口主要居住在辽东北部和中南部的 21 个村庄。其登记人口规模在 1768-1861 年间增长较快，从 1404 人增长到 3016 人。尽管在 1864 年，皇家内务府把该人群内一部分居民进行了重新划分，登记人口减少为 2381 人，但此后一直到 1909 年，人口又继续快速增长，1909 年达到 4378 人，整体增速比辽东北部其他地区人口增速高出许多。

地图 17 貂皮屯汉军人丁村庄

表 39 貂皮屯汉军人丁特征

人口地位	常规旗人
民族	汉、满族
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	无
地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	中等

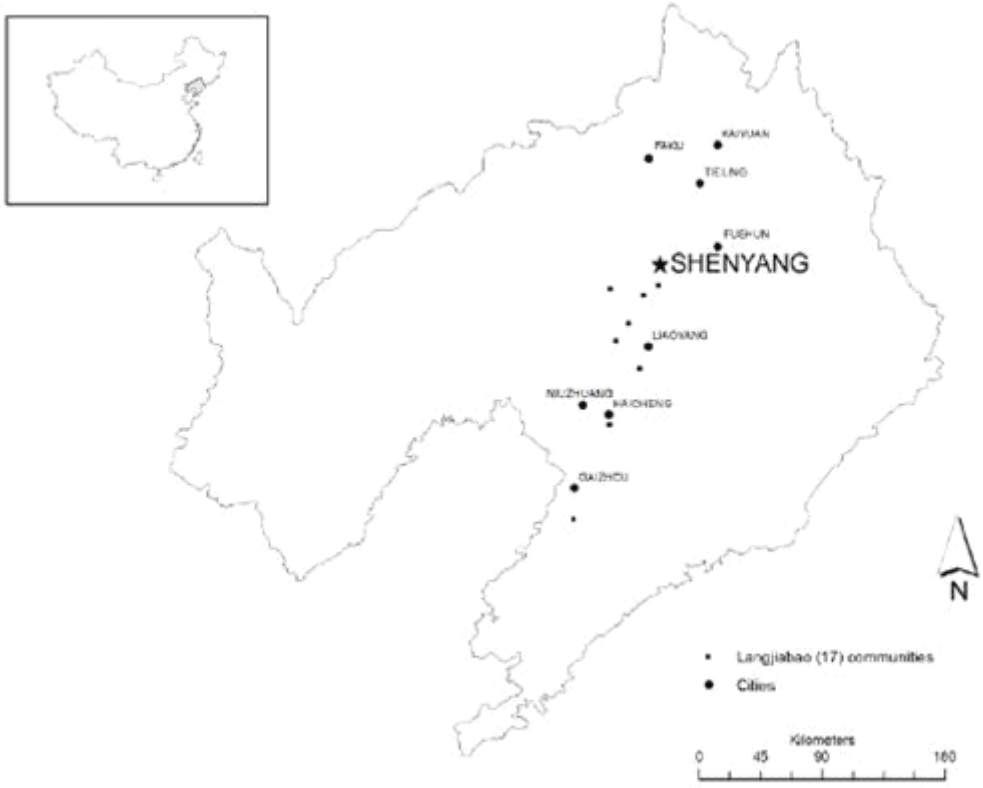
貂皮屯汉军人口属于皇家内务府下特殊义务旗人，其中男性多登记作为皮丁、染皮匠或者是猎人（shengding，dashengding）。意味着他们在国家提供的土地上生活的代价是需向皇家内务府提供特定的劳动产品（Ding, Guo, Lee, and Campbell 2004）。比如作为猎人，每位成年男性就需要上缴两只野猪和一些鹿肉，或者每年支付 15 两银子作为替代。

这一特殊义务人群主要集中居住在辽宁北部。表 21 中显示，99.89%的记录显示人口位于辽东北部，仅有 0.11%位于辽东中南部。鉴于该人口特殊义务旗人的身份，个人在地理层面和社会等级层面的流动性都非常有限。表 22 中，人口中仅有 11 位个人曾在辽宁两个地区生活过。同样，他们的从事国家提供的非农带薪岗位的机会也很稀少。如表 23 中总结的，在整个时间跨度里，仅有 0.74%的 18-60 岁男性的记录显示其拥有非农带薪岗位，没有一个人作为工匠。

从名字上看，尽管貂皮屯汉军人口的标题已说明人口主要为（汉军、汉族旗人），但其内还记录了个别蒙古族家庭的信息。1801-1816 年间户口册中没有出现过满族居民，1816 年之后，标记为满族居民的户数增长到三。这些满族人的情况一直保留在户口册中，直到 1858 年时他们的信息消失，不再编入登记。

6.Q郎家堡魏家屯汤岗子海州岳州汉军人丁(17)

郎家堡魏家屯汤岗子海州岳州汉军人丁数据库是由 1756-1909 年间的 25 卷郎家堡魏家屯汤岗子海州岳州汉军人丁户口册录入而来。数据库共包括 9092 位个人的 47331 次记录，这些人口主要登记居住在辽东中南部和中部地区的 27 个村庄。其人口规模的增长率属 CMGPD-LN 中最高的之一，从 1756 年的 713 人增长至 1909 年的 3496 人。



地图 18 郎家堡魏家屯汤岗子海州岳州汉军人丁村庄分布

表 40 郎家堡魏家屯汤岗子海州岳州汉军人丁特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	无
荣誉头衔的数量	19
是否可以参加科举考试	是
地理流动性	高
带俸禄的职位	中等

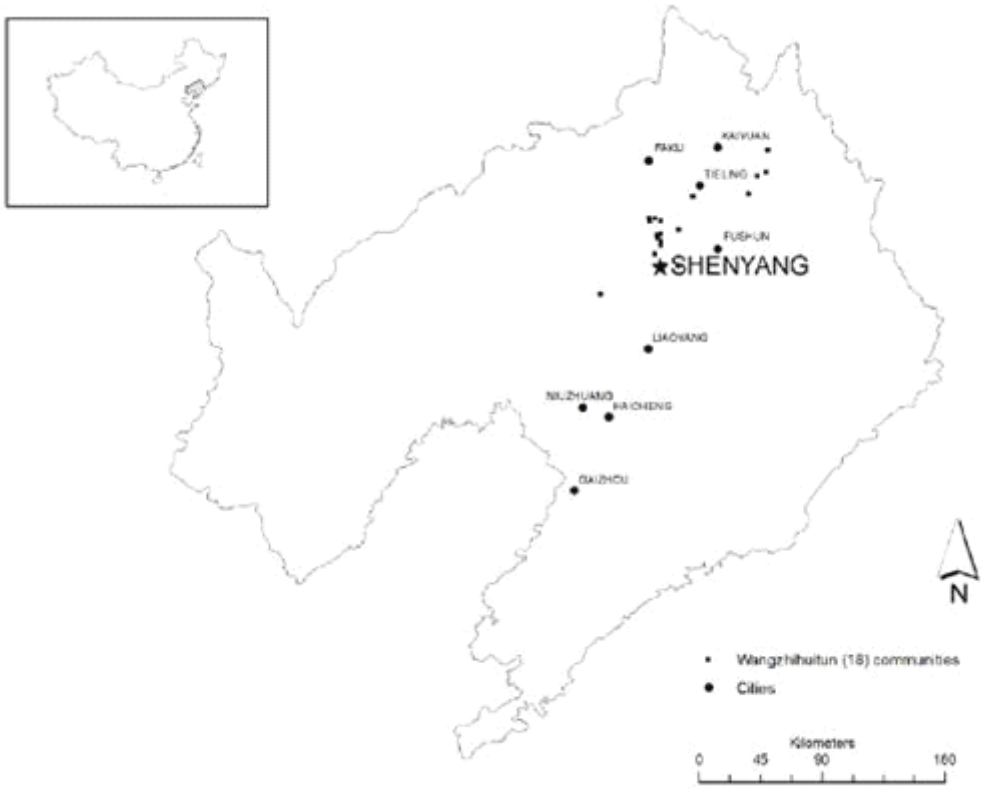
从此数据库的名称上包括了这么多个地名即可以看出，该人口的地理分布极为的分散，辽东的主要四个地区中都有其分布，其中居住在辽宁中南部的人口最多（66.48%的记录），辽宁北部最少（0.83%）。居住在中南部的人口主要是位于辽阳和牛庄等地，中部的人口则主要位于沈阳南郊，南部人口位于盖州和岳州。鉴于其人口居住如此分散，自然人口的地理流动性也较高。表 22 中显示，约有 8.3%登记的个人曾经在辽东的两个以上地区生活过，其中相当一部分人在三个地区生活过。

郎家堡魏家屯汤岗子海州岳州汉军人口大部分也属于普通旗人。仅有生活在辽宁中部和中南部的一部分人口中有特殊义务旗人甚至是庄园旗人的存在：生活在辽宁中部的 juliuhe 的一部分男性被称作“城庄头”或“城丁”，还有一些生活在牛庄地区的男性被注释为“瓦丁”。当然，从所占比例上讲，这些人口可以忽略。

虽然有上述特例的存在，郎家堡魏家屯汤岗子海州岳州人口从拥有国家提供非农带薪岗位角度看主要还是表现出作为普通旗人人群的一些特点。如表 23 中所示，1.38%的 18-60 岁男性记录显示其拥有非农带薪岗位，有 0.98%有军人的身份。但是，这些从事非农带薪岗位的机会的分布却是极为不均衡的，几乎所有的这些岗位都属于居住在辽宁中部的人口。整体来算，共有 4.15%的成年男性记录显示其拥有非农带薪岗位，而在其他三个地区的人口中仅有约 1%的男性拥有这类岗位。

6.R王指挥屯汉军人丁(18)

王指挥汉军人丁数据库由 1765-1909 年间的 28 卷王指挥屯汉军人丁户口册录入而来。数据库共包括了 10054 位个人的 60334 次记录，这些人主要居住在辽宁北部、中部和中南部的 44 个村庄。登记人口规模从 1765 年的 907 人增长至 1909 年的 2765 人。



地图 19 王指挥屯汉军人丁村庄分布

表 41 王指挥屯汉军人丁特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	网户达（只针对特殊特殊义务旗人）
行政长官	族长
军事长官	有
工匠	有
地方官员	无
荣誉头衔的数量	3
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	高

尽管王指挥屯本身就位于沈阳北郊，但事实上只有47.25%的登记人口居住在辽东中部，其他有36.18%生活在辽东北部，16.56%生活在辽东中南部，剩下还有很少一部分生活在辽

东南部（表 21）[55]。居住在辽东中部的人口主要集中于沈阳北郊和市内，生活在北部的人口主要位于铁岭地区，生活在中南部的人口主要集中在辽阳地区。该人口的地理流动性也较高，如表 22 中所示，有约 10%的登记人口曾经在辽东的两个地区以上生活过。

该人口不仅在地理上的分布较为分散，同样在人口组成上也同时包括普通旗人和特殊义务旗人。这些向皇家内务府提供特殊服务的人主要是渔民和牧民。除了这些特殊义务人口中的个人外，其余人口中 18-60 岁男性拥有非农带薪岗位的比例还是比较高（3.83%，表 23）。同时，人口中还有不少的工匠人口。这些拥有非农带薪岗位的人口主要集中在辽北的铁岭地区和中部的沈阳地区。如果把人口的地区分布考虑进去的话，辽北人口中有 3.77%，辽中人口中有 4.97%的成年男性从事非农带薪岗位，该比例超过了许多其他普通旗人人口。

6.5 阿尔吉善(19)

阿尔吉善数据库是由 1813-1909 年间的 20 卷阿尔吉善户口册录入而来。数据库共包括了 2517 人的 15098 次记录，这些人口主要生活在辽东中部和中南部的 14 个村庄。阿尔吉善户口册不仅在时间跨度上相对短，其所载人口规模也相对小且波动。户口册记录中显示，人口自 1813 年的 574 人增长到 1873 年的 837 人。1873 年后，其人口一直在 600-800 间浮动，最后到 1909 年的 755 人。

[55] 1765 年户口册中有 4 个观察值来自辽东南部。

地图 20 阿尔吉善村庄分布

表 42 阿尔吉善人口特征

人口地位	服务旗人
民族	汉
组织形式	根据父系族(一户或族)来组织但没有住宅家庭的信息
长官	管领
监管者	庄头
行政长官	庄头
军事长官	无
工匠	无
地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	否
地理流动性	非常高
带俸禄的职位	无

“阿尔吉善”很可能指 Songtu 的第二个儿子，所以这些人口也可能是指这位背叛康熙皇帝的叛徒的后代[56]。但至于“阿尔吉善”人口确切的身份目前还不清楚，他们与一般旗人人口制度的差异反映出其可能作是政治流放的庄园旗人，但也不同于后面将要介绍的五户和苏拜公人口。与一般的国有庄园旗人人口直接隶属于皇家内务府下某一旗不同的是，阿尔吉善人口是直接由控制皇家内务府下属国有庄园农场的会计司管理的[57]。根据户口册中反映的情况，这些人口由一位“庄头”进行总体管理。

表 21 中列出，53.7%的阿尔吉善人口登记记录显示个人居住在辽中的沈阳和 xingjing 等地，剩下 43%的人口则居住在辽东中南部的辽阳及附近地区。尽管从制度上讲，皇家内务府不允许庄园旗人人口随意迁移，但这一人口中似乎经常存在上述两个地区间的人口流动。表 22 中显示，有 23.2%的被登记个人都曾经在这两个地区生活过，这使得阿尔吉善人口成为了 CMGPD-LN 中所有人口中个人的地理流动性最大的部分。

鉴于阿尔吉善人口属于庄园旗人，自然其个人的社会经济地位比较低。首先表现在，很少有人有机会从事国家提供的非农带薪工作。表 23 中可见，在整个历史跨度期间，阿尔吉善人口中没有一个人获得从事非农岗位的机会。同样因为其社会经济地位不高，同普通旗人相比，所以人口登记的完整性也大打折扣。在整个数据库中，列出的 16 岁以下的女性的记录仅仅只有 23 条。

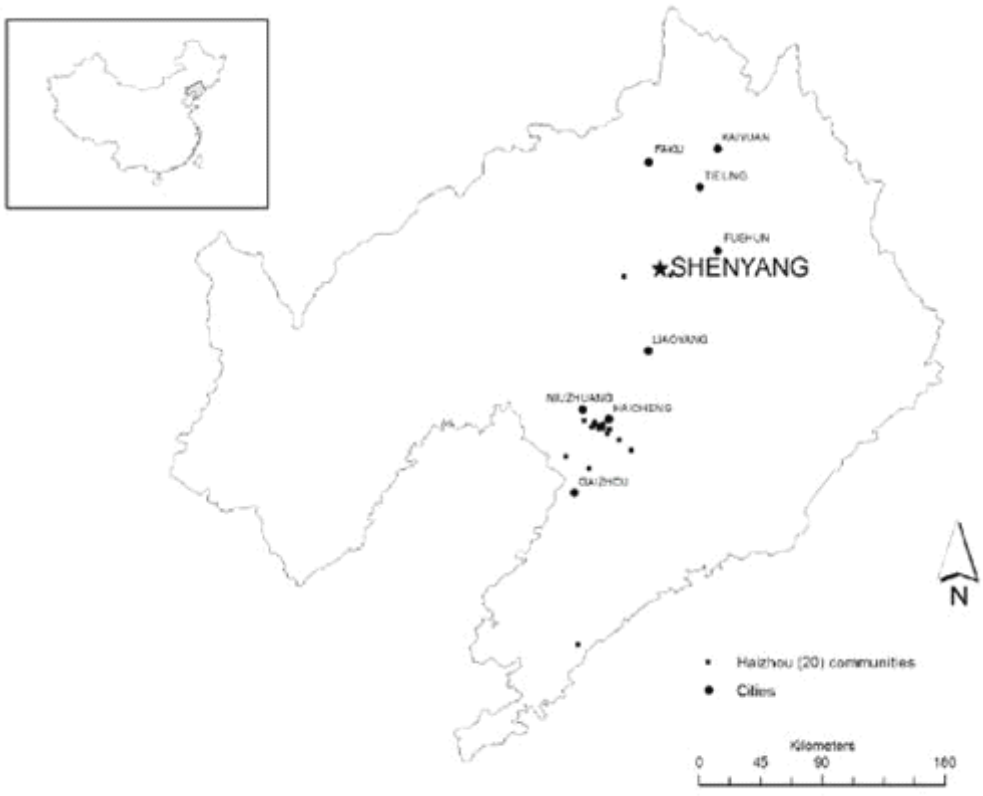
[56] 根据 Hummel （1943），Songotu（1636-1703)是康熙皇帝统治时期的一位官员。他是 Heseri clan 部落 Xiaocheng 的叔叔 ， 以及被任命在康熙皇帝在位期间协助他的摄政王 Sonin 的儿子,他也是太子胤礽的叔叔。根据 Oxnam1975 年的描述，Songotu 在协助康熙废黜另一位摄政王 Oboi 这件事上发挥了重要作用。 他曾是康熙最器重的一位大臣,但后来因贪污、叛国和弒君未遂被治罪，死在了监狱中。

[57] 内务府下有三支旗人：镶黄旗、正黄旗和正白旗。

从 1906 年始，阿尔吉善人的户口册的登记模式有所改变，这可能意味着其身份和地位有所改变。在最后两次登记，即 1906 和 1909 两次中，再没有出现监管者的信息，取而代之的是其个人居住地为“八旗”。有趣的是，这一改变却和发生在其他人口户口册中的变化相反——其他户口册中的“八旗”在这一时期内逐渐消失。

6.T海城汉军人丁(20)

海城汉军人丁数据库是由 1759-1909 年间的 26 卷海城汉军人丁户口册录入而来。数据库共包括 20032 位个人的 119186 次记录，这些人口主要生活在辽东所有四个地区：辽北、辽中、辽中南和辽南的 40 个村庄。这意味着海城汉军人丁是 CMGPD-LN 中所有 29 个人群中规模最大的。登记人口规模从 1759 年的 1596 人增长到 1909 年的 8950 人。



地图 21 海城汉军人丁村庄分布

表 43 海城汉军人丁特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领

监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	有
荣誉头衔的数量	47
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	中等

尽管海城汉军人口规模很大，但所载人口中女性子嗣的记录却非常至少，尤其是年轻的女孩子。在 1-5 岁的人口中，登记有 6488 位男孩而只有 154 位女孩。这种不平衡的情况直到 16-20 岁年龄段人群中才逐渐有所改善。

海城人口主要集中居住在辽宁中南部的牛庄地区（77.55%的记录），剩下有约 20%居住在辽宁南部的盖州地区。还有很少一部分人居住在辽宁中部和北部地区。海城汉军人口的地理流动性较高，表 22 中显示，有 12.5%的登记个人曾经在辽宁两个地区以上生活过。

海城汉军人口属于普通旗人。但 18-60 岁男性中仅仅只有 0.5%的个人拥有国家提供的非农带薪岗位（表 23），这其中的一些人是作为皇家内务府下属的商人（买卖人），可能比较富有。总体来看，该户口册中记录有许多荣誉性头衔——或购买、或受国家奖励——还有一些是通过国家科举考试而被授予的。这么多的荣誉头衔显示出，海城汉军人口很可能是 CMGPD-LN 中最富裕的人群之一。

6.U王多罗树满洲牲丁(21)

王多罗树满洲牲丁数据库是由 1864-1909 年间的 9 卷王多罗树满洲牲丁户口册录入而来。数据库共包括 2300 位个人的 9494 次记录。这些人口主要居住在辽北和辽中南的 10 个村庄。登记人口规模从 1864 年的 815 人增长至 1909 年的 1252 人。这部分户口册中几乎把女儿的信息完全忽略，仅仅有两例低于 16 岁的女性记录。

地图 22 王多罗树满洲牲丁村庄分布

表 44 王多罗树满洲牲丁特征

人口地位	常规旗人
民族	满族
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	朱显达
行政长官	族长
军事长官	无
工匠	无
地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	是
地理流动性	低
带俸禄的职位	中等

在所有户口册中，王多罗树满洲牲丁数据库的独有特色是：该人口系由少数名族组成。因为皇家内务府下属的人口大部分为汉族，而王多罗树牲丁人口则全部由满族个人组成，远多于八棵树和盖州两人口中的满族个人数。它是 CMGPD-LN 中包含有满足个人的 4 个数据库之一。

此外，根据其名称，我们还知道王多罗树满洲牲丁属于特殊义务旗人。他们在国家给予的土地上生活的同时最大的义务即需要向皇家内务府提供鹿肉产品。每位成年男性猎人（牲丁）需要向皇家内务府上缴 2 只野猪和 90 把的鹿肉。同样，他们也可以每年上缴 15 两白银作为替代。

作为一个特殊义务旗人人群，王多罗树满洲牲丁人口几乎没有地理上的流动性。表 21 显示，几乎所有的人口（99.48%的记录）生活在辽北的开原地区，仅仅有很少几户人家生活在辽中南的辽阳地区。此外，在登记的 45 年间，只有 4 位个人显示其在辽宁两个地区生活过。

虽然是作为特殊义务旗人，王多罗树满洲牲丁人口中的男性也有一部分的机会通过从事国家提供的非农带薪岗位来提高自己的社会地位。表 23 中，有 1.23%的 18-60 岁男性的记录显示其拥有非农带薪岗位，这些人主要是军人。

6.V王多罗树满洲人丁(22)

王多罗树满洲人丁数据库是由 1792-1909 年间的 20 卷王多罗树满洲人丁户口册录入而来。数据库共包括了 4899 位个人的 23233 次记录。登记人口规模从 1792 年的 703 人增长到 1909 年的 1439 人。

户口册中所记载的村庄，比如王多罗树，可能是位于现今辽北的开原地区。我们根据这个名字没有找到相关的村庄，但文献记载显示，现在的清河屯可能就是历史上的王多罗树，或者居住着他们的后代。我们在当地发现有很大的一个水库，所以我们推测原来的王多罗树可能是被淹入水中，而人口迁到了清河屯。

地图 23 王多罗树满洲人丁村庄分布

表 45 王多罗树满洲人丁特征

人口地位	常规旗人
民族	满族
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	无
地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	否
地理流动性	低
带俸禄的职位	高

由于在 117 年的时间段中仅留存了 20 卷户口册，其中一些较长的登记时间间隔给数据关联带来一定困难。里面最长的一次间隔是 1882-1903 年，这 21 年的间隔造成许多在早期登记的个人的人生信息不完整。

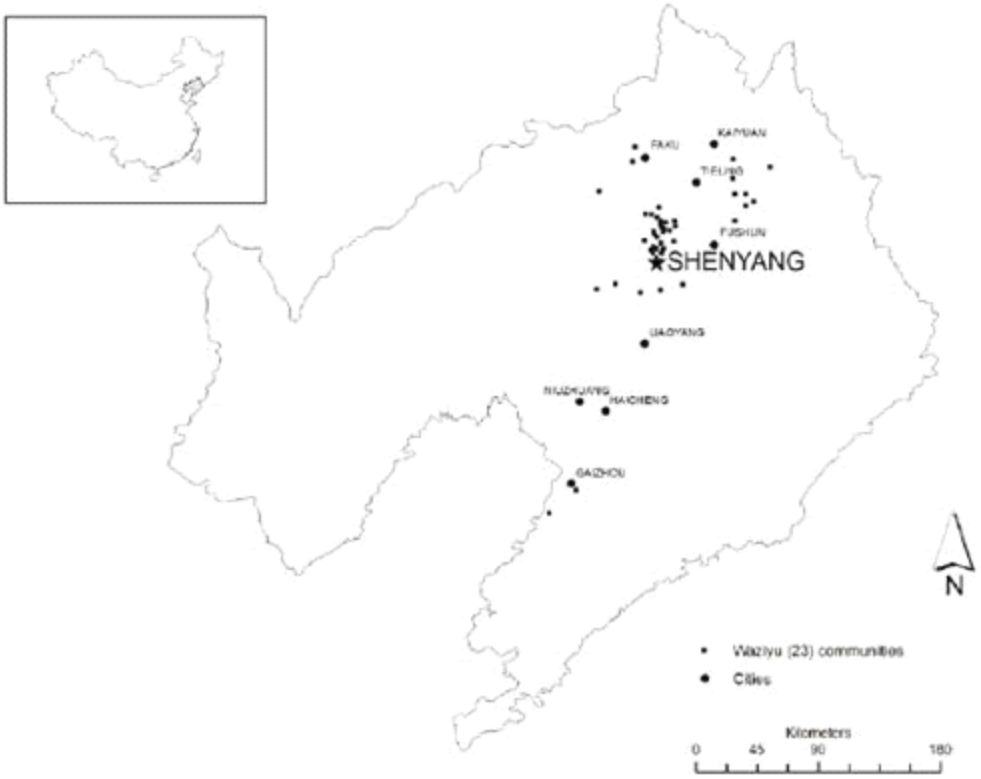
虽然王多罗树满洲人丁和王多罗树满洲牲丁人口在民族组成上一致，都全部为满族。但鉴于其不同的社会身份，王多罗树满洲人丁整体的社会地位较高，而牲丁作为特殊义务人群，自然地位较低。这种身份的差异可以从表中女儿信息的登记、人口类别以及他们所从事岗位的流动性即可以看出。

首先，和王多罗树满洲牲丁人口中非常有限的对于女儿的记录相比，王多罗树满洲人丁对 16 岁以下的女孩子的记录却是完整许多的。其中 1-5 岁人口性别比例是 275:100，远比其他一些人群该比例要小，事实上，仅有城内汉军人口中该比例才更低。这些显示出了王多罗树满洲人丁中的人们会更多的登记女儿的信息，其可能原因是在满族中女性的地位要比汉族中高，而对那些社会地位较高的满族家庭，他们的女儿有特别的机会可以去北京和沈阳进宫作为宫女。

第二，人口的组成以及从事相关岗位的机会上的差异也显示了王多罗树满洲人丁更高的社会地位。王多罗树满洲人丁主要是由一般国有农场和生产相关产品的专门国有农场的人口组成。其中绝大多数都生活在一般的国有农场中。然而，从 1825 年始，一些男性猎手（大牲丁和牲丁）开始出现在登记中。其后不久，又出现了牧民家庭（牧丁）。此外，该人口也有较多机会从事非农带薪岗位，其中，2.8%的 18-60 岁男性记录显示其拥有非农带薪岗位，主要为军人（表 23）。形成对比的是，其他三个包括有满族的人口——盖州满汉人丁、八棵树老古洞满汉人丁、王多罗树满洲牲丁——仅仅有 1.2%的男性记录显示其拥有非农带薪岗位，只相当于王多罗树满洲人丁该比例的一半。

6.W 瓦子峪汉军人丁(23)

瓦子峪汉军人丁数据库是由 1768-1909 年间的 27 卷瓦子峪汉军人丁户口册录入而来。数据库共包括 12884 位个人的 75042 次记录。这些人口主要居住在辽东北部、中部、中南部和南部的 58 个村庄。登记人口规模从 1768 年的 1569 增长至 1909 年的 3319 人。



地图 24 瓦子峪汉军人丁村庄分布

表 46 瓦子峪汉军人丁特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	有
工匠	有
地方官员	无
荣誉头衔的数量	23
是否可以参加科举考试	是

地理流动性	中等
带俸禄的职位	非常高

尽管瓦子峪汉军人口中大部分居住在辽宁中部地区，但人口整体的地理分布却较为分散。如表 21 中显示，约有 80%的人口记录来源于沈阳附近的村庄，属辽宁中部地区；剩下约 20%的人口则分散在其他三个地区。考虑到其地理分布的特点，该人口的地理流动性自然也较高。在表 22 中显示，有 15.14%的瓦子峪汉军人口曾经在辽宁两个以上地区生活过。

瓦子峪汉军人口也是由生活在一般旗地上的普通旗人和生活在特殊义务旗地上的特殊义务旗人群体组成。这些特殊义务旗人主要从事的工作包括打渔（渔丁）、放牧（牧长或牧丁）和种植棉花（棉丁）等等。此外，该户口册中还包括几户作为亲王府上仆人的居民[58]。抛开这些特殊义务人群不算，瓦子峪人口中其他的成年男性也有较多的机会从事非农带薪岗位。如表 23 中所示，5.36%的 18-60 岁男性记录中显示其拥有非农带薪岗位。如果把人口的地理分布考虑进去的话，辽东的北部、中部和南部地区成年男性中该比例超过 5%。而占成年男性比例 0.76%的工匠却主要集中在辽宁北部和中部地区。

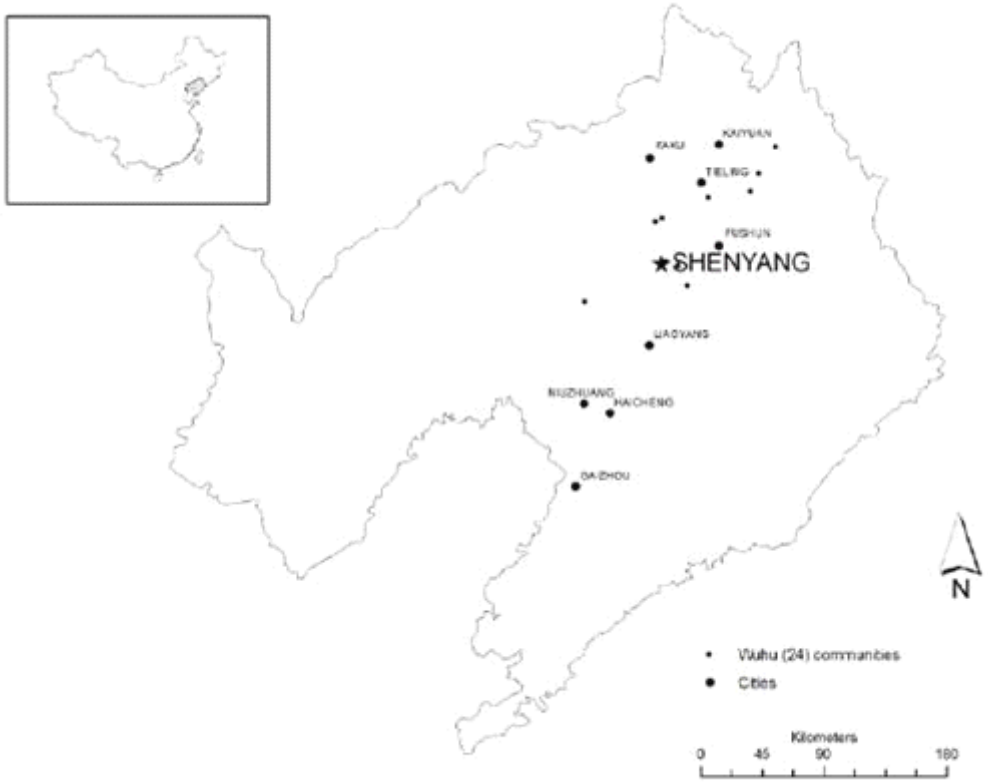
和其他许多户口册一样，1903 那一年的瓦子峪汉军户口册中形式也发生了变化。许多新的村庄名称出现在户口册中。这一居住地的变化也给数据关联造成了一定的麻烦，使得部分出现在之前户口册中的个人的人生信息不完整。我们推测这些村庄新名称的产生很可能和清末十年八旗制度的改革有关（Zheng1981）。

6.X五户汉军人丁(24)

五户汉军人丁数据库是由 27 卷五户汉军人丁户口册录入而来。这“五户”即指五个在和雍正皇帝（胤禛 1723-1735）有关皇位的政治斗争中被迫害并驱逐的贵族家族——Akina、Seshe、Yun’e、Sunu、Leshixin[59]。这些户口册产生的时间是 1789-1909 年间，共包括了 7687 位个人的 44790 次记录。这些人生活在辽东北部、中部和中南部地区的 62 个村庄。登记人口规模在 1789 年时为 1130 人，其后一直停留在此规模直到 1852 年，随后又继续增长到 1909 年的 3030 人。

[58] 这些家庭出现在 1780 和 1825 户口册中。

[59] 在这 5 个贵族中，Akina （Yunsi）是康熙皇帝(1662-1722)的第八个儿子。康熙皇帝死后，Akina 成为 Yinzhen 争夺皇位的对手。他试图在朝廷中组建一个政治利益团体，成员包括康熙皇帝的第九个儿子 Saisihei（Yuntang）、康熙皇帝的第十个儿子 Yun’e, Nurhachi 的曾曾孙 Sunu, 和 Sunu 的儿子 Leshixin。他们共同反对 Yinzhen 继位。Yinzhen 或者说雍正皇帝继位后，连续削掉了这五个贵族的权力和勋爵，剥夺了他们的财产，包括他们在辽东的地产（在 1726 年这些土地属于盛京内务府正黄旗下）。



地图 25 五户汉军人丁村庄分布

表 47 五户汉军人丁特征

人口地位	常规旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长
军事长官	无
工匠	无
地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	是
地理流动性	高
带俸禄的职位	中等

五户汉军人口的地理分布也较为分散。除了主要四个居住地区——辽中的沈阳、辽宁中南部的辽阳、广宁和牛庄以外，还有部分人口生活在辽北的开原、铁岭和辽中、辽中南的其他一些地区。如表 21 中所示，有约 50%人口登记为居住在辽宁中南部，有 24%生活在辽北，剩下 25%生活在辽中。鉴于如此分散的人口分布，五户汉军人口的地理流动性也较高。 表

22 中显示，有 16.64%的个人曾经在两个以上地区生活过。

尽管五户汉军人口中包括了一些在庄园进行劳动的被流放贵族成员，但其仍旧属于普通旗人人群。整个五户人口直接受正黄旗的管辖，而所有的居民则归族长管理，这些标志即是普通旗人所不同于被奴役人口的区别。

从五户人口中所从事的职业种类便可以了解到清政府对不同政治流放人群的待遇不同。和 26 号数据库——苏拜公人口一样，五户人口也有资格从事国家的非农带薪岗位。如表 23 中列出的，有 1.22%的 18-60 岁男性记录显示其拥有非农带薪岗位，主要为当地的收税人（领催）。这就和阿尔吉善人口的情形不同——他们中没有人拥有这类岗位。不过，和苏拜公人丁一样，五户人口中也没有出现工匠和军人的身份，而这两种才是在绝大多数普通旗人中最常见的。这种既非奴役人口又非普通的有完全特权的旗人的情况，也反映出庄园旗人人口中多样化的本质组成。

6.Y棉花盐靛庄头庄丁(25)

棉花盐靛庄头庄丁人口包括了多种特殊义务人群：种植棉花、制盐、染匠等。这一数据库是由 1756-1909 年间的 29 卷棉花盐靛庄头庄丁户口册录入而来。该数据库共包括了 13404 位个人的 76738 次记录，这些人口主要分布在辽北、辽中、辽中南以及辽南的 83 个村庄中。

地图 26 棉花盐靛庄头庄丁村庄分布

表 48 棉花盐靛庄头庄丁特征

人口地位	常规旗人
民族	汉、高丽、高丽满族
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	庄头
行政长官	族长
军事长官	无
工匠	无
地方官员	无
荣誉头衔的数量	3
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	低

尽管人口总体上的地理分布较为分散，但实际上同一种特殊义务人口的居住还是相对集中的。种植棉花的人口主要位于辽中南的牛庄地区和辽南的盖州地区。制盐的人口则主要位于辽南的牛庄地区，那里沿海、故适宜制盐。染匠人口居住在辽中的沈阳地区和辽北的铁岭地区。表 21 中显示，12.2%的记录对应人口生活在辽北地区，10.6%生活在辽中地区，42.7%生活在辽中南地区，最后 34.6%则生活在辽南地区。

在整个时间跨度中，棉花盐靛庄头庄丁人口的登记经历了几次变化，造成其登记人口规模也相应有波动。人口从 1756 年的 2170 人增长到 1855 年的 2854 人。1867 年，八旗管理层把其中一部分人口划走并重新登记为棉花盐靛小门人丁户口册[60]。对于这种重新划分的原因，我们推测可能是去掉了那些 bonded nature 特别低的人群，因为这里的“小门”字面意思即社会底层的小户人家。因此，1867 年登记的人口数跌至 1851 人。而在 1873 年，所有的染匠人口也被划出，但这倒并没有长久的影响到人口的规模，到 1909 年人口又增长至 4067 人。鉴于其人口的分布较分散，加之几次再管理归属上的变化，人口整体的地理流动性较高。表 22 中显示，有 14%的登记个人曾经在辽东两个地区以上生活过。

由于人口属于特殊义务人群，故其中男性个人职位上的流动性也非常有限，从拥有非农带薪岗位的情况便可以看出。表 23 中列出，仅仅有 53 例（0.2%）的 18-60 岁男性记录显示其从事非农带薪岗位。而像工匠或士兵这类身份，人口中一个也没有。

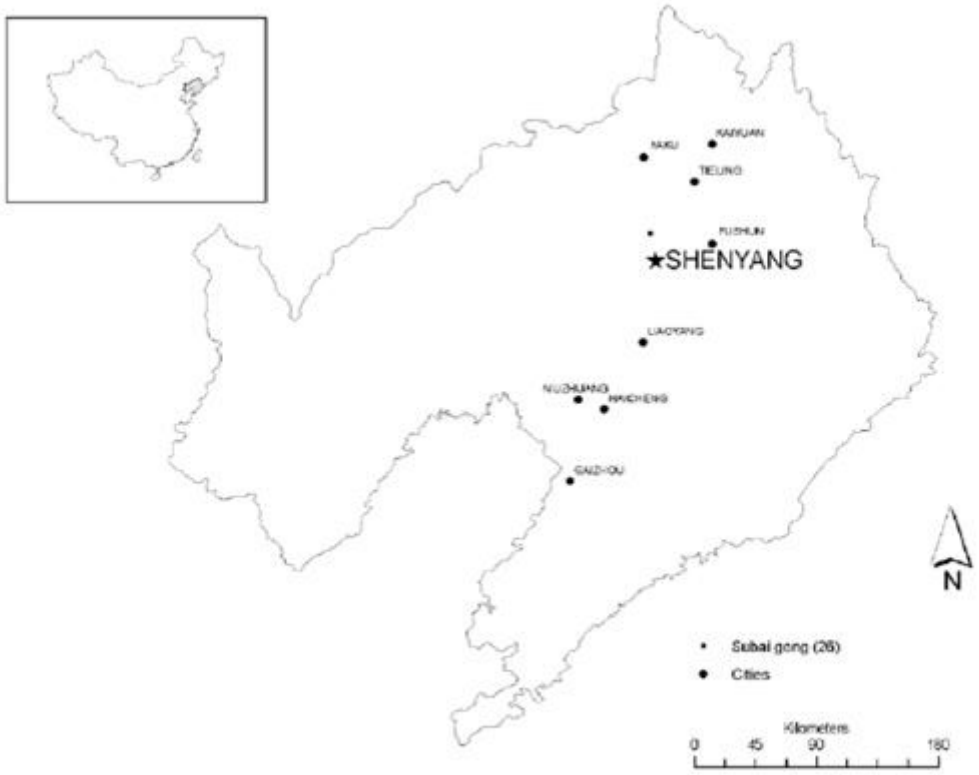
还有一点有趣的是，该人口的民族组成是非常多样的。除了汉族旗人外，还包括一些高丽人和高丽满洲家庭。高丽人口规模自一开始就比较小，随着时间推移又逐渐萎缩，1756 年有 5 户高丽居民，1768 年 3 户，到 1909 年仅仅有 1 户剩下。相反的是，在十九世纪末期，盖州满汉人口（8 号数据库）中的高丽满洲居民却增长快速。1864 年时仅仅只有 1 户高丽居民，到 1867 年新增加了 46 户，随后该数目一直维持到 1903 年，高丽满洲居民又减少到 18

[60] 这些独立的户口册变成了 CMGPD-LN 的 29 个数据库。我们将会分别描述这些人口。

户，到 1909 年，仍旧还剩下 15 户。

6.2 苏拜公人丁(26)

苏拜公人丁数据库是由 1864-1909 年间的 10 卷苏拜公人丁户口册录入而来。该人口属于特殊义务人口，其中成年男性从事的主要工作即是看守坟墓（坟丁）。所有 10 卷户口册中，除了 1864 年那卷，其余卷中所有的村庄都有注释“苏拜公坟所属各屯”即指苏拜公坟下属各村庄。这些人口很可能就是清朝开国将军苏拜公坟墓的看守人[61]。



地图 27 苏拜公人丁村庄分布

表 49 苏拜公人丁特征

人口地位	附属旗人（1830 年之前）常规旗人（1830 年之后）
民族	蒙古族
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	无
行政长官	族长

[61] Subai 的父亲是 Nurhachi（满洲国和清朝的开国皇帝）早期的同事之一。Subai 和他的哥哥 Wubai 成为满洲军队的重要将领，帮助皇帝占领中原并打败了吴三桂。作为回报，Subai 被封了爵。

军事长官	无
工匠	无
地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	无
地理流动性	高
带俸禄的职位	低

该数据库所涉及的时间跨度较短，包含人口也较少。共有来自于辽东北部、中部、中南部的四个村庄的 879 位个人的 4162 次记录。苏拜公户口册中缺失绝大部分女孩儿的记录，整个数据库中 16 岁以下女孩儿的记录只有 22 例。

根据 1864 年的户口册（图 5.1）封面上的说明，苏拜公人口属于蒙古普通旗人（正身蒙古）。然而，他们的普通旗人的身份却是在 1830 年才确定下来的（查明根基），这就意味着在那之前，他们的身份仍然是个未知数。此外，这一说明还指出其中的成年男性在 1860 年以前是没有机会参加国家科举考试的。即使是在苏拜公人口的社会地位提高之后，他们所从事的职业依然有很大的限制性——比如仍旧不能成为官员或士兵。这些关于苏拜公人口的社会状况的描述向我们揭示了，即使从十九世纪开始，特殊义务旗人的社会地位有所提高，但和普通旗人相比依旧有差距。

从人口所从事的职业种类看，也支持上述关于职业受限制的论断。他们几乎没有从事国家提供非农带薪岗位的机会。从表 23 中看，仅仅有 3 例 18-60 岁男性拥有这种职位，而工匠和军人则是一个都没有。

鉴于其人口主要义务为看守坟墓，所以绝大多数都居住在沈阳附近——也是苏拜的坟墓最可能的位置。如表 21 显示，65.6%的人口记录显示其居住在辽中的沈阳地区，有 13.4%居住在辽北，还有 21%则生活在辽中南部。该人口的地理流动性却是非常之高，表 22 中列出，在仅仅 45 年的时间段中，有 16%的登记个人曾经在辽东两个地区生活过。

再一点即是，苏拜公人丁是我们找到的皇家内务府下属人群中唯一的蒙古族人口。

图 16　1864 年苏拜公人丁户口册封面

6.AA 开档(27)

开档数据库是由 1810-1909 年间的 12 卷开档户口册录入而来，所谓“开档”是八旗制度中一个特殊的词，指那些新登记的、从属性质的特殊人口，和投降的人口（投从）、被抱养的人口（抱养）类似，通常这三类人群也都划分在一起。开档人口因为属于特殊从属人群，自然其社会经济地位都很低下。该数据库共包括有来自辽东北部、中部、中南部的 14 个村庄的 2717 位个人的 8835 次记录。

地图 28 开档村庄分布

表 50 开档人口特征

人口地位	附属旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	窝棚头（针对一部分人）
行政长官	族长
军事长官	无
工匠	有

地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	否
地理流动性	低
带俸禄的职位	中等

开档人口的组成在 20 世纪初发生了很大的变化，不仅是分布还是人口规模都受到了影响。该人群 1810 年时人口为 628 人，其中绝大多数居住在辽中的沈阳周围。只有约 25%的记录位于辽东北部和中南部地区。此后，人口一直维持在这种水平直到 1903 年，一批 612 位来自于辽北的成员加入其中。使得人员集中的重心转到了辽东北部地区。这一趋势一直延续到 1906 年，在那一年，又有 355 位来自辽北的个人加入进来，人口达到了 1018 人。1909 年登记时，人口为 1193 人。

开档人口中人的地理流动性较一般，因此可以知道地理上分布的变化不是由人口迁移产生，而是因为登记人口的变化造成的结果。如表 22 所示，仅仅只有 8.6%的个人曾经在两地地区生活过。

作为特殊附属人群，开档户口册所载人口争取非农带薪岗位时有很大的限制。表 23 中列出，只有 39 例 18-60 岁男性记录显示其有非农带薪岗位；每一年间，约只有 3 到 8 位成年男性可以得到这类位置。此外，这些位置中也从来没有军人或官员，而全部属于工匠，并主要集中在辽中地区。另外，开档人口中还有一些被指定为皇家内务府从事其他义务的人，他们主要是种植棉花和采蜜。

6.AB 开档投从抱养人丁(28)

开档投从抱养数据库是由 1792-1909 年间的 25 卷开档投从（投降）抱养（被收养）人丁户口册录入而来。数据库共包括 5087 位个人的 30137 次记录，这些人口主要居住在辽北、辽中、辽南和辽中南的 49 个村庄。

地图 29 开档投从抱养人丁村庄分布

表 51 开档投从抱养人丁特征

人口地位	附属旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	庄头（针对一部分人）
行政长官	族长
军事长官	无
工匠	无
地方官员	无
荣誉头衔的数量	0
是否可以参加科举考试	否
地理流动性	中等
带俸禄的职位	低

和开档人口类似，开档投从抱养人口也属于那类社会底层的特殊从属人群。在 1864 年之前，该户口册主要分为两部分：每一村庄的开档投从抱养人口和由苏拜墓管辖的村庄中的人口。从 1864 年开始，所有登记在苏拜墓下属的人口被重新划分入一个新的类别——苏拜公人丁户口册。我们从苏拜公人口长久以来一直归属其中便可以看出他们也一直被认为属于

社会底层人群。

户口册所包括人口的种类的变化自然也影响到了人口的规模。1792 年登记人口为 1090 人，1861 年增长至 1205 人。1864 年苏拜公人口撤出后，总数回落到 887 人。随后人口逐渐增长到 1909 年的 1659 人。此外，开档投从抱养人口的地理分布也很广，如表 21 中总结，33.4%的记录位于辽东北部，15.7%的记录位于辽东中部，33.7%记录位于辽东中南部，最后 17.2%的人口记录位于辽东南部。

从人口从事职业的种类也可瞥见从属人口的很低的社会地位。在整个人口中，仅仅有一户有非农带薪的岗位，还是属于底层的“领催”。其余的人口主要从事拾棉花、制盐、守墓和庄丁等工作。他们不仅需要提供相应的特殊劳动义务，还有监管人员（庄头）来监督他们的劳动过程。

尽管该人口社会地位低，但地理上的流动性却不低。如表 22 中列出的，有超过 16%的开档投从抱养人口曾经在辽东两个以上地区生活过。这一地理流动性甚至超过了绝大多数普通旗人人口。

6.AC 棉花盐靛小门人丁(29)

棉花盐靛小门人丁数据库是由 1861-1909 年间的 10 卷棉花盐靛小门人丁户口册录入而来。这一系列户口册属于 25 号数据库棉花盐靛庄头庄丁的附属户口册。最初，该人口的记录是棉花盐靛庄头庄丁户口册的一部分。但从 1861 年开始，皇家内务府重新把这部分人口划分为“小门”——即社会地位低，并在小门户口册中单独登记。

地图 30 棉花盐靛小门人丁村庄特征

表 52 棉花盐靛小门人丁特征

人口地位	附属旗人
民族	汉
组织形式	根据另户和父系族(一户或族)来组织
长官	佐领
监管者	庄头（针对一部分人）
行政长官	族长
军事长官	无
工匠	无
地方官员	无
荣誉头衔的数量	2
是否可以参加科举考试	是
地理流动性	中等
带俸禄的职位	无

该数据库共包括 5277 位个人的 25900 次记录，这些人主要居住在辽北、辽中、辽中南和辽南的 44 个村庄。在 5277 位个人中，有 1266 位在 25 号和 29 号数据库中均有记录。我们把这些 1266 位个人在小门户口册中的记录和其在棉花盐靛庄头庄丁户口册的原始记录进行了关联。小门人口规模从 1861 年的 1488 人增长至 1909 年的 3871 人。

和棉花盐靛庄头庄丁人口以及其他附属人群一样，小门人口的地理分布也较广。如表 21 中显示，绝大多数记录显示人口居住在辽东中南部的牛庄（43.5%）和辽东南部的盖州地区（42.7%）。此外，还有 7.4%居住在辽北地区，6.3%居住在辽中的沈阳地区。还有一点和棉花盐靛庄头庄丁人口类似的是，小门人口的地理流动性也很高，如表 22 中显示的，有 14%的登记个人曾经在辽东两个地区生活过。

小门人口从事职业的种类也和棉花盐靛庄头庄丁类似，或者反映出更底层的社会地位。小门人口没有任何机会从事非农带薪岗位。如表 23 中显示，小门人口中没有一个人拥有该类职位。相反，他们还需要向皇家内务府提供特殊劳动义务：拾棉花、制盐和染布。

小门户口册中对于女儿信息的忽略还要远比棉花盐靛庄头庄丁人口更甚。棉花盐靛庄头庄丁人口中 1-15 岁个人的性别比为 668:100，已经非常之高。而小门人口中极少记录女儿的信息，全部户口册中仅仅有 23 例，这使得为一些实际目的来计算棉花盐靛庄头庄丁人口的性别比例已经失效了。